

## SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

### Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

### STAFF USE ONLY

Date completed: 05-13-03  
Searcher: Beverly C4994  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 23  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

#### Search Site

\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

#### Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

#### Vendors

\_\_\_\_\_ IG Suite  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
✓ Other CGN

**THIS PAGE BLANK (USPTO)**

Region

```

FT FT /note-"This region is 22% identical to histone
FT FT macro-H2A.1 protein"
FT FT Region 319..485.
FT FT /note-"This region is 26% identical to histone
FT FT macro-H2A.1 protein"
FT FT Modified-site 392..395
FT FT /note-"Tyrosine phosphorylation site"
FT FT Modified-site 495..498
FT FT /note-"Tyrosine phosphorylation site"
FT FT Region 508..709
FT FT /note-"partially homologous to protein families,
FT FT myosin heavy chain and cytoskeleton linkers
FT FT ezrin-radixin-moesin (ERM)"
FT FT Domain 781..786
FT FT /label=Proline-rich domain
PA PA WOZ00026231-A1.
XX XX
XX XX 11-MAY-2000.
XX XX
XX XX 29-OCT-1999; 99W0-US25439.
XX XX PF 29-OCT-1998; 98U5-0106383.
XX XX PR 29-OCT-1998; 98US-0106448.
XX XX PR 30-OCT-1998; 98US-0106448.
XX XX (DAND ) DANA FARBER CANCER INST INC.
PA PA
XX XX Shipp M, Aguilar R, Yakushiji Y;
XX XX WPI: 2000-365563/31.
XX XX DR N-PSDB: AAD00629.
XX XX
PT PT New B-aggressive lymphoma nucleic acid for identifying cells exhibiting
PT PT or predisposed to malignancies such as lymphoma, sarcoma, carcinoma and
PT PT melanoma
XX XX
XX XX Claim 12; Fig 1; 151pp; English.
CC CC The present sequence is a B-aggressive lymphoma (BAL).
CC CC protein obtained by screening two human cDNA libraries derived from
CC CC anti-immunoglobulin activated splenocytes and the Raji Burkitts lymphoma
CC CC cell line cloned into pCM8. The BAL locus is mapped to chromosome 3q21.
CC CC BAL was identified to be differentially expressed in DLB-CL (diffuse
CC CC large B-cell lymphoma), the most common non-Hodgkin's lymphoma, by
CC CC differential display technique. It was found to be more abundant in
CC CC tumours from patients with high risk fatal DLB-CL disease than low risk
CC CC cured disease. BAL transcripts are most abundant in lymphoid organs e.g.
CC CC spleen, foetal liver and peripheral blood, and several non-haematopoietic
CC CC organs e.g. heart and skeletal muscle. BAL is involved in modulation of
CC CC cellular adhesion and aggressiveness/severity of malignancy such as
CC CC DLB-CL. BAL nucleic acid and protein are useful for identifying cells
CC CC exhibiting or predisposed to malignancies such as lymphoma, sarcoma,
CC CC carcinoma and myeloma. BAL proteins are useful as immunogens to raise
CC CC anti-BAL antibodies. BAL modulators or BAL molecules can be used for
CC CC prophylactic and therapeutic treatment of a subject susceptible to or
CC CC having a disorder associated with aberrant BAL expression or activity,
CC CC such as non-Hodgkin's lymphoma. It is also useful in vaccine preparation
XX XX
XX XX Sequence 854 AA:
SO SO
Query Match 100.0%; Score 4447; DB 21; Length 854;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY 1 MDFSMVGAAGAAVYNKSCRITSLILPKVAQAFPMWRKGNEECPLPYKCSEFGALGENY 60
Db 1 MDFSMVGAAGAAVYNKSCRITSLILPKVAQAFPMWRKGNEECPLPYKCSEFGALGENY 60
OY 61 SWQIPINHNDFIKLNNEBOLCEYLQKKFCSICITLVSPOEGNSKSLSQYFRKMILPRLTEL 120
Db 61 SWQIPINHNDFIKLNNEBOLCEYLQKKFCSICITLVSPOEGNSKSLSQYFRKMILPRLTEL 120
OY 121 SVWKDDLTTHAVDAVVAANEDLLHGCGIALALVKACGEFIODESKOFVARYGKVSAEII 180

```

|          |          |   |      |
|----------|----------|---|------|
| Db       | 121      | SVWKDDLTTHNAVDAVVAANDDLHGGLALALVKKAGCFEIOESKQFVARVYGVSA               | 180  |
| QY       | 181      | AVTAGGRLPCKQITIHAVGPRMMEKDGCGCTGKIORAIVSLNLYITKNTHIKTVAIPALS          | 240  |
| Db       | 181      | AVTAGGRLPCKQITIHAGPRMMEKDGCGCTGKIORAIVSLNLYITKNTHIKTVAIPALS           | 240  |
| QY       | 241      | SGIQFPLNLCTKTIVETIRVSLQGRPMMSLKEIHLVSNEDPTVAAFKAASEPILGKSE            | 3000 |
| Db       | 241      | SGIQFPLNLCTKTIVETIRVSLQGRPMMSLKEIHLVSNEDPTVAAFKAASEPILGKSE            | 3000 |
| QY       | 301      | LGOETTPSFNAVNNLTLOITVOGHIEMQPADVIVNSVNHDTITVGVAKSILQOAGVEM            | 3600 |
| Db       | 301      | LGOETTPSFNAVNNLTLOITVOGHIEMQPADVIVNSVNHDTITVGVAKSILQOAGVEM            | 3600 |
| QY       | 361      | KSEFLAKAKQFOFSOLVLYTKGNELCCKTIYHVLHMSSEPKPILKHNKCELEKCTEQ             | 4200 |
| Db       | 361      | KSEFLAKAKQFOFSOLVLYTKGNELCCKTIYHVLHMSSEPKPILKHNKCELEKCTEQ             | 4200 |
| QY       | 421      | NITSISFPALGTGSMETKKEETAELFDEVLTFAKHVKKHQLTVKFIPTDLEIYKAFS             | 4800 |
| Db       | 421      | NITSISFPALGTGSMETKKEETAELFDEVLTFAKHVKKHQLTVKFIPTDLEIYKAFS             | 4800 |
| QY       | 481      | SEMAKRSKMISLNNYSVPQSTREKRENGLEARSAPAINLGFNVEEYEAHAMIQRTLSL            | 5400 |
| Db       | 481      | SEMAKRSKMISLNNYSVPQSTREKRENGLEARSAPAINLGFNVEEYEAHAMIQRTLSL            | 5400 |
| QY       | 541      | QNHIIENNNHLYLGRKEHNDLSOLQKTSVYSITEIISPGRTELEIGARADILEVYMI             | 6000 |
| Db       | 541      | QNHIIENNNHLYLGRKEHNDLSOLQKTSVYSITEIISPGRTELEIGARADILEVYMI             | 6000 |
| QY       | 601      | EDMLCKYOEEMARKKEKGLMRSIGOMTLOQKQKODEKKENIIFLKCVPRTOELLDKKQ            | 6600 |
| Db       | 601      | EDMLCKYOEEMARKKEKGLMRSIGOMTLOQKQKODEKKENIIFLKCVPRTOELLDKKQ            | 6600 |
| QY       | 661      | FEKCGLOVLKYEKIDNEVULMAAFQRRKKRMEBEKLRQPVSHRLFEQVYVQFCNVYCVGF          | 7200 |
| Db       | 661      | FEKCGLOVLKYEKIDNEVULMAAFQRRKKRMEBEKLRQPVSHRLFEQVYVQFCNVYCVGF          | 7200 |
| QY       | 721      | QRMTSTCDPRYGGAGIYFTKMLKYLAKKAKKISADKLIVYFEAEVLTGFCQGHPLNIY            | 7800 |
| Db       | 721      | QRMTSTCDPRYGGAGIYFTKMLKYLAKKAKKISADKLIVYFEAEVLTGFCQGHPLNIY            | 7800 |
| QY       | 781      | PPPLSPGALIDGHDVNVNVSSEPFVYFSGMOALPQYLMTCQTOFYVQSOYSSGPMRPA            | 8400 |
| Db       | 781      | PPPLSPGALIDGHDVNVNVSSEPFVYFSGMOALPQYLMTCQTOFYVQSOYSSGPMRPA            | 8400 |
| QY       | 841      | QHWRGFGASGSPVD 854  |      |
| Db       | 841      | QHWRGFGASGSPVD 854  |      |
| RESULT 2 |          |   |      |
| ABB05392 | ABB05392 | standard; Protein: 819 AA.  |      |
| AC       | AC       | ABB05392;   |      |
| XX       | XX       | 10-Apr-2002 (first entry)   |      |
| DT       | DT       |   |      |
| DE       | DE       | Human interferon-induced tetraspan (IT) protein #2.                   |      |
| XX       | XX       |   |      |
| KM       | KM       | Human; interferon-induced tetraspan; IT; cytosolic; anti-HIV;         |      |
| KW       | KW       | immunosuppressive; antiinflammatory gene therapy; autoimmune disease; |      |
| KM       | KM       | inflammatory disease; cardiovascular disease; metabolic disease;      |      |
| KW       | KW       | cancer growth inhibition; metastasis.                                 |      |
| OS       | OS       | Homo sapiens.   |      |
| XX       | XX       |   |      |
| FT       | FT       | Key Location/Qualifiers   |      |
| FT       | FT       | Modified-site 24..26  |      |
| FT       | FT       | /note="asn is N-glycosylated"   |      |
| FT       | FT       | Modified-site 69..71  |      |

```

FT FT /note= "Asn is N-glycosylated"
FT FT 198..217
FT FT /label= "TM"
FT FT /note= "transmembrane domain"
FT FT 218..269
FT FT /note= "intracellular loop"
FT FT 270..291
FT FT /label= "TM"
FT FT /note= "transmembrane domain"
FT FT 292..730
FT FT /note= "large extracellular loop"
FT FT 386..388
FT FT /note= "Asn is N-glycosylated"
FT FT 459..461
FT FT /note= "Asn is N-glycosylated"
FT FT 731..754
FT FT /label= "TM"
FT FT /note= "transmembrane domain"
FT FT 755..768
FT FT /note= "intracellular loop"
FT FT 769..788
FT FT /label= "TM"
FT FT /note= "transmembrane domain"
PN W0200196399-A2.
PD 20-DEC-2001.
PE 15-JUN-2001; 2001WO-US19305.
PR 15-JUN-2000; 20000US-211565P.
XX
XX (CURA-) CURAGEN CORP.
PA (BIO) BIOGEN INC.
XX
PI Peyman JA, Da Silva A, Hockman P;
XX
XX WPI: 2002-098056/13.
DR N-PSDB: ABA93022.
XX
XX Novel isolated interferon-induced tetraspan polypeptide, useful for
PT treating autoimmune, inflammatory, cardiovascular and metabolic
PT diseases, and for inhibiting cancer growth and metastasis -
XX
XX Claim 1; Fig 1; 105pp; English.
XX
XX The present sequence represents a human interferon-induced tetraspan
CC (IT) protein (I). (I) has cytosolic, anti-HIV, immunosuppressive and
CC antiinflammatory activities. The polynucleotide (II) encoding (I) can be
CC used in gene therapy. (I) and (II) can be used for treating or preventing
CC a pathology associated with (I) in a subject, preferably human, or for
CC the manufacture of a medicament for treating a syndrome associated with a
CC human disease selected from a pathology associated with (I), (I) and (II)
CC can be used for treating autoimmune diseases, inflammatory diseases,
CC cardiovascular diseases, metabolic diseases, and for inhibiting cancer
CC growth and metastasis. (I) or (II) can be used as diagnostic and/or
CC prognostic markers, in gene therapy, as research tools, and for tissue
CC regeneration in vitro and in vivo.
XX
XX Sequence 819 AA:

```

```

Query Match 94.6%; Score 4208.5; DB 23; Length 819;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 815; Conservative 3; Mismatches 1; Indels 35; Gaps 1;

```

```

QY 1 MDESMVAGAAAYNKRSGITSLSLFQVFAQIRPQMKGTGTECLPKCCEGTALGENTY 60
DB 1 MDESMVAGAAAYNKR-----SETGALGENTY 25
QY 61 SMQIPINHNDRKILKNNRQLEVLONKFGCISTLSPVQGNKSLQVFRKNLTPREL 120
DB 26 SMQIPINHNDRKILKNNRQLEVLONKFGCISTLSPVQGNKSLQVFRKNLTPREL 85

```

```

QY 121 SWKRDLLTTHAVDAVNNANEDLLHGGGLALALYKAGFGEIOESKOFVARYGKVSAGEI 180
DB 86 SWKRDLLTTHAVDAVNNANEDLLHGGGLALALYKAGFGEIOESKOFVARYGKVSAGEI 145
QY 181 AVTGAGRLPCAKQIHAVGPWMEMDKQCTGKLOBAIVSLNVIYKTKHTKIVAIPALS 240
DB 146 AVTGAGRLPCAKQIHAVGPWMEMDKQCTGKLOBAIVSLNVIYKTKHTKIVAIPALS 205
QY 241 SGIFQFPLNLCTKTIVETIRVSLQCKPMMSNIKEIHLVSNEDPYTAAKKASEFLGKSE 300
DB 206 SGIFQFPLNLCTKTIVETIRVSLQCKPMMSNIKEIHLVSNEDPYTAAKKASEFLGKSE 265
QY 301 LGOETTPSFNMAVNNLLOIYQGHIEWQTADVIVNSNPNHDITGPAKSLIOQAGEM 360
DB 266 LGOETTPSFNMAVNNLLOIYQGHIEWQTADVIVNSNPNHDITGPAKSLIOQAGEM 325
QY 361 KSEFLATKAKQFQRSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCEIQ 420
DB 326 KSEFLATKAKQFQRSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCEIQ 385
QY 421 NITTSFPLGIGNMEIKKETAELTLPEDVLTFAADHYKHQUTYFVIFPDDLEYKAFS 480
DB 386 NITTSFPLGIGNMEIKKETAELTLPEDVLTFAADHYKHQUTYFVIFPDDLEYKAFS 445
QY 481 SEMAKRSKMLSLNNYSVPQSTREEKRENGLEARSPAIIMLGMFVBNEMEYAHAMIORISL 540
DB 446 SEMAKRSKMLSLNNYSVPQSTREEKRENGLEARSPAIIMLGMFVBNEMEYAHAMIORISL 505
QY 541 QNHIIENNNHILYGRKHDILISQLOKTSVSITEIISPGRTLEIEGARADLIEVNNI 600
DB 506 QNHIIENNNHILYGRKHDILISQLOKTSVSITEIISPGRTLEIEGARADLIEVNNI 565
QY 601 EDMLCVOEEMARKKERGLMSLGOWTIOQKTODEMKNIEIFLCPYPPOELIDOKKO 660
DB 566 EDMLCVOEEMARKKERGLMSLGOWTIOQKTODEMKNIEIFLCPYPPOELIDOKKO 625
QY 661 FEKCGLOVLKYEKIDNEVLMFAFORKKRMEBEKILRQPVSHRLFQOVYQFCNVVCYRGF 720
DB 626 FEKCGLOVLKYEKIDNEVLMFAFORKKRMEBEKILRQPVSHRLFQOVYQFCNVVCYRGF 685
QY 721 QRMVSTPCDPRYAGIYFTKLNKILAEKAKKISADKLIYFEAVLTGFCQGHPLNIV 780
DB 686 QRMVSTPCDPRYAGIYFTKLNKILAEKAKKISADKLIYFEAVLTGFCQGHPLNIV 745
QY 781 PPLSPGALDGHDSVDVNVSPETVFIRSGMAIROYLMTGTOEYVOSODYSSGPMRPA 840
DB 746 PPLSPGALDGHDSVDVNVSPETVFIRSGMAIROYLMTGTOEYVOSODYSSGPMRPA 805
QY 841 QHPWRGFAAGSPVD 854
DB 806 QHPWRGFAAGSPVD 819

```

```

RESULT 3
ABB05391
ID ABB05391 standard; Protein: 780 AA.
AC ABB05391;

```

```

DT 10-APR-2002 (first entry)
XX
XX
XX

```

```

DE Human interferon-induced tetraspan (IT) protein #1.
XX
XX

```

```

KW Human: interferon-induced tetraspan; IT; cytosolic; anti-HIV;
KW immunosuppressive; antiinflammatory gene therapy; autoimmune disease;
KW inflammatory disease; cardiovascular disease; metabolic disease;
KW cancer growth inhibition; metastasis.
XX
XX

```

```

OS Homo sapiens.
XX
XX

```

```

FH Key Location/Qualifiers
FT Misc-difference 417..419
FT /note= "encoded by ACATT"

```

| FT | Misc-difference 770..772  | /note= "encoded by A0ATTT" |
|----|---|----------------------------|
| XX | WO200196399-A2.   |                            |
| XX |   |                            |
| XX | 20-DEC-2001.  |                            |
| XX |   |                            |
| XX | 15-JUN-2001; 2001WO-US19305.  |                            |
| XX |   |                            |
| XX | 15-JUN-2000; 2000US-211565P.  |                            |
| XX |   |                            |
| XX | (CURA-) CURAGEN CORP.   |                            |
| XX | (BIOL) BIOGEN INC.  |                            |
| XX |   |                            |
| XX | Peyman JA, Da Silva A, Hockman P;   |                            |
| XX |   |                            |
| XX | WPI; 2002-098056/13.  |                            |
| XX | N-PSDB; ABA93022.   |                            |
| XX |   |                            |
| XX | Novel isolated interferon-induced tetraspan polypeptide, useful for       |                            |
| XX | treating autoimmune, inflammatory, cardiovascular and metabolic           |                            |
| XX | diseases, and for inhibiting cancer growth and metastasis -               |                            |
| XX |   |                            |
| XX | Claim 1; Page 8; 105pp; English.  |                            |
| XX |   |                            |
| XX | The present sequence represents a human interferon-induced tetraspan      |                            |
| XX | (IIT) protein (I). (I) has cytosolic, anti-HIV, immunosuppressive and     |                            |
| XX | antiinflammatory activities. The polynucleotide (II) encoding (I) can be  |                            |
| XX | used in gene therapy. (I) and (II) can be used for treating or preventing |                            |
| XX | a pathology associated with (I) in a subject, preferably human, or for    |                            |
| XX | the manufacture of a medicament for treating a syndrome associated with a |                            |
| XX | human disease selected from a pathology associated with (I). (I) and (II) |                            |
| XX | can be used for treating autoimmune diseases, inflammatory diseases,      |                            |
| XX | cardiovascular diseases, metabolic diseases, and for inhibiting cancer    |                            |
| XX | growth and metastasis. (I) or (II) can be used as diagnostic and/or       |                            |
| XX | prognostic markers, in gene therapy, as research tools, and for tissue    |                            |
| XX | regeneration in vitro and in vivo.  |                            |
| XX |   |                            |
| XX | Sequence 780 AA:  |                            |
| XX |   |                            |
| XX | Query Match 88.3%; Score 3927.5; DB 23; Length 780;                       |                            |
| XX | Best Local Similarity 94.5%; Pred. No. 0;                                 |                            |
| XX | Matches 770; Conservative 3; Mismatches 5; Indels 37; Gaps 3.             |                            |
| QY | 1 MDESVAGAAANENSGRTSLSLFQKFAQIFQMKRGNTPECLPYKSGTGAAGENY 60                |                            |
| DB | 1 MDESVAGAAANENK-----SETGALGENY 25  |                            |
| QY | 61 SMOIPIINHDEFILKNNEHOLCEVLONKFGCISTLWSPVOEGNSKSLAFYRKMLPRIEL 120        |                            |
| DB | 26 SMOIPIINHDEFILKNNEHOLCEVLONKFGCISTLWSPVOEGNSKSLAFYRKMLPRIEL 85         |                            |
| QY | 121 SYWKDGLTHAVDAVYVNAANEDLHGGLALALVYKAGGFEIOEESKQFVARYKVSAGEI 180        |                            |
| DB | 86 SYWKDGLTHAVDAVYVNAANEDLHGGLALALVYKAGGFEIOEESKQFVARYKVSAGEI 145         |                            |
| QY | 181 AVTGAGRPLCKOIIHAVGPRMHEMPOKOGTGLOPRAIVSLWVYIKNHIKTVAIPALS 240         |                            |
| DB | 146 AVTGAGRPLCKOIIHAVGPRMHEMPOKOGTGLOPRAIVSLWVYIKNHIKTVAIPALS 205         |                            |
| QY | 241 SGIFQFPLNLCKTIVETIRVSLQGRPMKSNLKEIHLVSNEDPTVAAFKASEFTLGKSE 300        |                            |
| DB | 206 SGIFQFPLNLCKTIVETIRVSLQGRPMKSNLKEIHLVSNEDPTVAAFKASEFTLGKSE 265        |                            |
| QY | 301 LGOETTPSPNNAVNNLTIQIVQGHIMQVOTADVIYNSVPHITVGPVAKSLIOQAGVEM 360        |                            |
| DB | 266 LGOETTPSPNNAVNNLTIQIVQGHIMQVOTADVIYNSVPHITVGPVAKSLIOQAGVEM 325        |                            |
| QY | 361 KSEFLATKAKQQRQSOLVLTGKFNLFECYIYHVLHMSFFPKPOLLIHMKCELEKICIO 420        |                            |
| DB | 326 KSEFLATKAKQQRQSOLVLTGKFNLFECYIYHVLHMSFFPKPOLLIHMKCELEKICIO 385        |                            |
| QY | 421 NITSISFPALGTGNEIKETAAILDFEVLTE-AKDHHKQDLTVKFEVITPDLTEYKAF 479         |                            |

|          |          |   |     |
|----------|----------|---|-----|
| Db       | 386      | NITISFPAIGCTGMMELKKEETAELIFDEVLLITAKDHVKQLIVKVIPTDLEITKAF               | 445 |
| Oy       | 480      | SESEAKRSKMLSLNNYSVPDSTREKKRENGLEARSPAINLMGFVNEEYEAHANIORILS             | 539 |
| Db       | 446      | SESEAKRSKMLSLNNYSVPDSTREKKRENGLEARSPAINLMGFVNEEYEAHANIORILS             | 505 |
| Oy       | 540      | LONHHIENNHHIYLKREKHEHDILSOLOKSSVSTREIETISPORTELELEGARADIEVNM            | 599 |
| Db       | 506      | LONHHIENNHHIYLKREKHEHDILSOLOKSSVSTREIETISPORTELELEGARADIEVNM            | 565 |
| Oy       | 600      | IEDMLCKVQDEMAKKERGLMSLQWTTQQOKTODEKKNIIFLKCPVPPTOELLDOKK                | 659 |
| Db       | 566      | IEDMLCKVQDEMAKKERGLMSLQWTTQQOKTODEKKNIIFLKCPVPPTOELLDOKK                | 625 |
| Oy       | 660      | QFEKCGLOVLKVKKIDINEVYMAFQKKKKMEKXLRQPVSHRLPFOQVPQFCVNCRVG               | 719 |
| Db       | 626      | QFEKCGLOVLKVKKIDINEVYMAFQKKKKMEKXLRQPVSHRLPFOQVPQFCVNCRVG               | 685 |
| Oy       | 720      | FORMYSTPCDPKYGAGIYFTKMLKLNLAERAKKISADRLIYVFEAEVLTGFFCGHPPLNT            | 779 |
| Db       | 686      | FORMYSTPCDPKYGAGIYFTKMLKLNLAERAKKISADRLIYVFEAEVLTGFFCGHPPLNT            | 745 |
| Oy       | 780      | VPPPLSPGALDGHDSVDNVNSSPEFF-VFFSGMOA                                     | 813 |
| Db       | 746      | VPPPLSPGALDGHDSVDNVNSSPEIITVIFSGMOA                                     | 780 |
| RESULT 4 |          |   |     |
| ABB50253 |          |   |     |
| ID       | ABB50253 | standard; Protein; 653 AA.  |     |
| XX       | AC       | ABB50253;   |     |
| XX       | D7       | 05-FEB-2002 (first entry)   |     |
| DE       |          | Human transcription factor TRFX-104.                                    |     |
| XX       | KW       | Human: transcription factor: TRFX; cell proliferative disease;          |     |
| KW       | KW       | autoimmune disease; inflammation; neurological disease;                 |     |
| KW       | KW       | developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV;  |     |
| XX       | OS       | neuroprotective; antiinflammatory; gene therapy.                        |     |
| XX       | XX       | Homo sapiens.   |     |
| XX       | XX       | MO200172777-A2.   |     |
| XX       | XX       | 04-OCT-2001.  |     |
| XX       | XX       | 13-MAR-2001; 2001MO-US08117.  |     |
| XX       | XX       | 13-MAR-2000; 2000US-0188986.  |     |
| XX       | XX       | (INCYTE GENOMICS INC.   |     |
| XX       | XX       | Hillman JL, Baughn MR, Yue H, Lal P, Lu DM, Patterson C;                |     |
| XX       | XX       | Azizmaiz Y, Bandman O, Tang YT, Mathur P, Shah P, Au-Young J;           |     |
| XX       | XX       | Reddy R;  |     |
| XX       | XX       | WPI: 2001-570896/64.  |     |
| XX       | XX       | N-PSDB: ABA63077.   |     |
| XX       | XX       | Novel transcription factor polypeptides, used to treat diseases         |     |
| XX       | XX       | associated with altered activity and expression of TRFX, and to screen  |     |
| XX       | XX       | for agents capable of modulating its activity.                          |     |
| XX       | XX       | Claim 1: Pages 247-249; 327pp; English.                                 |     |
| XX       | XX       | The present sequence is the protein sequence for a human transcription  |     |
| XX       | XX       | factor. The transcription factor and its coding sequence are useful in  |     |
| XX       | XX       | the diagnosis, treatment and prevention of diseases associated with     |     |
| XX       | XX       | altered expression of the transcription factor e.g. cell proliferative, |     |
| XX       | XX       | autoimmune/inflammatory, neurological and developmental disorders. A    |     |

CC number of specific disorders/diseases are given in the specification,  
 CC including: arteriosclerosis, cirrhosis, hepatitis, cancers, AIDS,  
 CC allergies, anaemia, asthma, autoimmune thyroiditis, bronchitis, atopic  
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,  
 CC Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis,  
 CC psoriasis, rheumatoid arthritis, systemic lupus erythematosus, ulcerative  
 CC colitis, uveitis, Alzheimer's disease, Huntington's disease, Parkinson's  
 CC disease, stroke, and viral, bacterial, fungal and protozoal infections.  
 XX  
 SQ Sequence 653 AA;

Query Match 76.2%; Score 3388; DB 22; Length 653;  
 Best Local Similarity 99.8%; Pred. No. 1,2e-278;  
 Matches 652; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 202 MEMKOGCTGLORAIVSLNVIYKNTHTKTAIPALSSGIFOPPLNCTIVETTRV 261  
 DB 1 MEMKOGCTGLORAIVSLNVIYKNTHTKTAIPALSSGIFOPPLNCTIVETTRV 60  
 QY 262 SLQGRPMNSNLKEIHLVSNEDPTVAAPKASEPILGKSELQETTPSFNAVNNLTLOI 321  
 DB 61 SLQGRPMNSNLKEIHLVSNEDPTVAAPKASEPILGKSELQETTPSFNAVNNLTLOI 120  
 QY 322 VQGIEMQTAIVYNSVPHDITVGPVAKSLIOAGVAKSEPLATKAKQORSOLVLT 381  
 DB 121 VQGIEMQTAIVYNSVPHDITVGPVAKSLIOAGVAKSEPLATKAKQORSOLVLT 180  
 QY 382 KGFNLFCYIYHVLMHSFEPKPOLIKHAMKCLEKICFONITSTISFPLGCGNMEIKET 441  
 DB 181 KGFNLFCYIYHVLMHSFEPKPOLIKHAMKCLEKICFONITSTISFPLGCGNMEIKET 240  
 QY 442 AAEILFDEVLTFARDVHKQLTVKVFIEPTDLEIYKAFSSPMARKSKMLSNYSVPOST 501  
 DB 241 AAEILFDEVLTFARDVHKQLTVKVFIEPTDLEIYKAFSSPMARKSKMLSNYSVPOST 300  
 QY 502 REERRENGLEKSPAINLMGFVNEEMTAHAMIQRIISLONNHIIENNHIITLYGKRENDI 561  
 DB 301 REERRENGLEKSPAINLMGFVNEEMTAHAMIQRIISLONNHIIENNHIITLYGKRENDI 360  
 QY 562 LSQLOKTSVSVITELISGRTLEIEGARADLIEVNMIEMLCKVQEMARKKREGIMR 621  
 DB 361 LSQLOKTSVSVITELISGRTLEIEGARADLIEVNMIEMLCKVQEMARKKREGIMR 420  
 QY 622 SLGQMTIOQOKTODEMKNIIIFLCPVPPTDLDQKKQFEGCGLVAKVEKIDNEVMA 681  
 DB 421 SLGQMTIOQOKTODEMKNIIIFLCPVPPTDLDQKKQFEGCGLVAKVEKIDNEVMA 480  
 QY 682 AFQRRKKMEKELHROPVSHRLEFQOVYQFCNVVCRVGFQRMVSTPCDPKYGAGIYFTKN 741  
 DB 481 AFQRRKKMEKELHROPVSHRLEFQOVYQFCNVVCRVGFQRMVSTPCDPKYGAGIYFTKN 540  
 QY 742 LKNIAEKAKKISADKLIVFEAEVLTFGFCQGHPLNIVPPPLSGALDGHDSVVDNVS 801  
 DB 541 LKNIAEKAKKISADKLIVFEAEVLTFGFCQGHPLNIVPPPLSGALDGHDSVVDNVS 600  
 QY 802 PETEVIFSGMOAIPOLYMTCTQEVYOSODYSSGPRPFAQHPMRGFAAGSPVD 854  
 DB 601 PETEVIFSGMOAIPOLYMTCTQEVYOSODYSSGPRPFAQHPMRGFAAGSPVD 653

RESULT 5  
 AAY71083  
 ID AAY71083 standard; Protein: 826 AA.  
 XX  
 AC AAY71083;  
 XX  
 XX 29-AUG-2000 (first entry)  
 DT  
 XX  
 DE Mouse B-aggressive lymphoma (BAL) protein-short form.  
 XX  
 KW Mouse: B-aggressive lymphoma; BAL; tumour; malignancy;  
 KW differential expression; DLB-CL; Diffuse large B-cell lymphoma;  
 KW cytosolic; vaccine; treatment; diagnosis; non-Hodgkin's lymphoma;

KW cellular adhesion; sarcoma; carcinoma; myeloma.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 748..753  
 FT /label= Proline-rich\_domain  
 FT Region 74..250  
 FT /note="This region is 24% identical to histone  
 FT macro-H2A.1 protein"  
 FT 261..450  
 FT /note="This region is 26% identical to histone  
 FT macro-H2A.1 protein"  
 FT  
 PD 11-MAY-2000.  
 XX  
 XX 29-OCT-1999; 99WD-US25439.  
 XX  
 PR 29-OCT-1998; 98US-0106383.  
 PR 30-OCT-1998; 98US-0106448.  
 XX  
 PA (DAND ) DNA FARMER CANCER INST INC.  
 PI Shipp M, Aguilar R, Yakushijin Y;  
 XX  
 DR WPI: 2000-365563/31.  
 DR N-PSDB: MAD00630.  
 XX  
 PT New B-aggressive lymphoma nucleic acid for identifying cells exhibiting  
 PT or predisposed to malignancies such as lymphoma, sarcoma, carcinoma and  
 PT myeloma  
 PT  
 PS Claim 12; Fig 2: 15pp; English.  
 XX  
 XX The present sequence is a short form of B-aggressive  
 CC lymphoma (BAL) protein obtained from mouse (Balb-c) spleen.  
 CC BAL was identified to be differentially expressed in DLB-CL (diffuse  
 CC large B-cell lymphoma), the most common non-Hodgkin's lymphoma, by  
 CC differential display technique. It was found to be more abundant in  
 CC tumours from patients with high risk fatal DLB-CL disease than low risk  
 CC cured disease. BAL transcripts are most abundant in lymphoid organs e.g.  
 CC spleen, foetal liver and peripheral blood, and several non-haematopoietic  
 CC organs e.g. heart and skeletal muscle. BAL is involved in modulation of  
 CC cellular adhesion and aggressiveness/severity of malignancy such as  
 CC DLB-CL. BAL nucleic acid and protein are useful for identifying cells  
 CC exhibiting or predisposed to malignancies such as lymphoma, sarcoma,  
 CC carcinoma and myeloma. BAL proteins are useful as immunogens to raise  
 CC anti-BAL antibodies. BAL modulators or BAL molecules can be used for  
 CC prophylactic and therapeutic treatment of a subject susceptible to or  
 CC having a disorder associated with aberrant BAL expression or activity,  
 CC such as non-Hodgkin's lymphoma. It is also useful in vaccine preparation.  
 XX  
 SQ Sequence 826 AA;

Query Match 55.2%; Score 2453; DB 21; Length 826;  
 Best Local Similarity 61.4%; Pred. No. 4.5e-199;  
 Matches 508; Conservative 96; Mismatches 206; Indels 18; Gaps 7;

QY 41 NTECLPYKCSFEGALGENYSWQIPINHNDFKILKNRROLCFVLONKFGCISTLVSPVQ 100  
 DB 2 DTWMAAPARERANNLSLEHYRMOQPIKINVFELIKSNSQICEVLOKNGFCISTLVSCPL 61  
 QY 101 EGNKSLO-VERKMLTPRIELSVWKDILTTAAVDAVVAANEDLHGGLALAYKAGCF 159  
 DB 62 AGSSSPAORVFRRTLIPIELISVWKDILTRHVYDAVVAANEDLHGGLAGLSTVKTGCF 121  
 QY 160 EIOESKOPVARYGVSGELAVTAGRLPCKQIIHAVGPRMMDKGCCTKIORATVS 219  
 DB 122 EIOESKRIIANVSKISVGGIVITGAGRLPCHLIHAAGPRVTVYNSOTALELFAIRN 181  
 QY 220 ILNVIYKNTHTKTAIPALSSGIFOPPLNCTIVETTRVSLQGRPMNSNLKEIHLVS 279

```

182 ILDVTVYDKIRKIVAIIPALSSGIFOPPLDCTSIILETIRLRYFOQDKOMEGNLEIHVLS 241
QY 280 NEDPTVAAPFASPIILGK--SELGOETTP--SFNMAVNVNLTLOIVOGHIEQOTADVI 334
Db 242 NEDPTVAAPFASPIILGK--SELGOETTP--SFNMAVNVNLTLOIVOGHIEQOTADVI 301
QY 335 VNSVNPHTITGVPAKSILOOAGVEKSEFLATRAKOFORSQVLVTKGFNLFCRYTYHV 394
Db 302 GNSGYMDPFKSGRAVOSILRQAGVEMEKEL--DKVNLSTDOEVAVWIKGFKLSGOYFHV 359
QY 395 LHWSEFPKPOLTKHAKMECELEKTEONITSPALGTGNNETKEKTAELTPEVLTFA 454
Db 360 AMHSQIMKYOILKDAKMSCKLEKCPDINSISFALGTGLDLKSTPAQIMFEVFAFA 419
QY 455 KDHYKHQTLKVFVIFPTDLEIYKAFSEMAKRSKMLSNYS--VPOSTREKRENGLE 511
Db 420 KEHEKRLTKVIVIFPVDEVEYKIFVAMKTKRSNEMLSGSGALALQWSSGDEORGL 479
QY 512 ARSPAINLMGFVNEEMYEANAMTIRISLONHHIENNHIILYGRKEDHILSLOKTSSV 571
Db 480 AGSPAINLMGFVNEEMYEANAMTIRISLONHHIENNHIILYGRKEDHILSLOKTSSV 539
QY 572 STEFIIISPGRELEIEGARADLIEVNMIEDMLCKVOEAMARKERGLMRSLOGMTIOOQ 631
Db 540 SISETVSPRATLEIKRPOADLDAVMKIECMCDVOEBAVAGKREKMLMSLSGGQNOQE 599
QY 632 KTDENKENTIFLKCPRPTOELLDOKKOFEKCGLOVLKVEKIDNEVJMAAFQKRRKME 691
Db 600 KL-DKMEESTYFORYPASLQELDRKKOFKFCGLWVVOYQIDNKYLLAFOKRRKME 658
QY 692 EKLHROVSHRLFOQVYVORCNVYVGRVGFQRMYSTPCDPKYGAGITTKNLKLAERAKK 751
Db 659 ERPRKGGSSORLFOQVYVORCNVYVGRVGFQRMYSTPCDPKYGAGITTKNLKLAERAKK 718
QY 752 ISADKLIYFEAEVLTGFCOGHPLNIVPPPLSPGALIDGDSVVDVNSPETFVIRSGM 811
Db 719 TTSRDKLIYFEAEVLTGFCOGHPLNIVPPPLSPGALIDGDSVVDVNSPETFVIRSGM 778
QY 812 QALPQYLTMTCTOXY-----VQSDYSSGPMRPAQHPRMGFASGSPV 853
Db 779 QAMPLYLMTCTODRTFQSHPMMSQDYSYSGPMVSSLSQSEWVLNGSSV 826

```

```

RESULT 6
AAM79380
ID AAM79380 standard; Protein: 360 AA.
XX
AC AAM79380;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3026.
XX
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.

```

```

PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR N-PSDB; AAK52513.
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy
XX
PS Claim 20; Page 233; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78396-AAK60302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 360 AA;

```

```

Query Match 42.5%; Score 1888; DB 22; Length 360;
Best Local Similarity 100.0%; Pred. No. 1.2e-151;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 497 VPOSTREKRENGLEARSPAINLMGFVNEEMYEANAMTIRISLONHHIENNHIILYGR 556
Db 3 VPOSTREKRENGLEARSPAINLMGFVNEEMYEANAMTIRISLONHHIENNHIILYGR 62
QY 557 KEHDILSLOKTSSVSTTEIISPGRELEIEGARADLIEVNMIEDMLCKVOEAMARKKE 616
Db 63 KEHDILSLOKTSSVSTTEIISPGRELEIEGARADLIEVNMIEDMLCKVOEAMARKKE 122
QY 617 RGLMRSLOGMTIOOQKQODEKKEIIFLKCPRPTOELLDOKKOFEKCGLOVLKVEKIDN 676
Db 123 RGLMRSLOGMTIOOQKQODEKKEIIFLKCPRPTOELLDOKKOFEKCGLOVLKVEKIDN 182
QY 677 EVLMAAFQRRKKMEELKHQPVSHRLFOQVYVORCNVYVGRVGFQRMYSTPCDPKYGAGI 736
Db 183 EVLMAAFQRRKKMEELKHQPVSHRLFOQVYVORCNVYVGRVGFQRMYSTPCDPKYGAGI 242
QY 737 YFRKLNKLAEKAKKISAAKLIYFEAEVLTGFCOGHPLNIVPPPLSPGALIDGHSV 796
Db 243 YFRKLNKLAEKAKKISAAKLIYFEAEVLTGFCOGHPLNIVPPPLSPGALIDGHSV 302
QY 797 DNVSSEPTFVIFSGMQAIPQYLTMTCTOXYVQSDYSSGPMRPAQHPRMGFASGSPVD 854
Db 303 DNVSSEPTFVIFSGMQAIPQYLTMTCTOXYVQSDYSSGPMRPAQHPRMGFASGSPVD 360

```

```

RESULT 7
AAM78396
ID AAM78396 standard; Protein: 335 AA.
XX
AC AAM78396;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1058.
XX

```



KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xie AJ, Yang Y, Wejman T, Goodrich R;  
 DR WPI: 2001-476283/51.  
 DR N-PSDB; AAK51529.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS Claim 20; Page 3291; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78333-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 CC  
 XX  
 SQ Sequence 335 AA;

Query Match 39.7%; Score 1765; DB 22; Length 335;  
 Best Local Similarity 99.7%; Pred. No. 3.1e-141;  
 Matches 334; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 520 MGFVNEEYEAHMTORILSLONHHIENNHILYGRKEHDLISOLQTSSTVSTIEIISP 579  
 DB 1 MGFVNEEYEAHMTORILSLONHHIENNHILYGRKEHDLISOLQTSSTVSTIEIISP 60  
 QY 580 GRTELEIEGARADILEVVMNIEDMLCKVOEEMARKKERGLMRSLOQWTIOOKTODEMKE 639  
 DB 61 GRTELEIEGARADILEVVMNIEDMLCKVOEEMARKKERGLMRSLOQWTIOOKTODEMKE 120  
 QY 640 NIILKCPVPPTOELDLQKQFEKCGLOLVKVERIDNEVLAARQRRKKHMEKLRQPV 699  
 DB 121 NIILKCPVPPTOELDLQKQFEKCGLOLVKVERIDNEVLAARQRRKKHMEKLRQPV 180  
 QY 700 SHRLFOOVYQFCNVVCGVGFQRMVSTPCDPKYGAGIYFTNKLKLAARKKISADKLI 759  
 DB 181 SHRLFOOVYQFCNVVCGVGFQRMVSTPCDPKYGAGIYFTNKLKLAARKKISADKLI 240

QY 760 VYFEAEVLTGFECGHPPLNIVPPLSPALDGHDSVVDNVSSPETFVIFSGMAIPOYLW 819  
 DB 241 VYFEAEVLTGFECGHPPLNIVPPLSPALDGHDSVVDNVSSPETFVIFSGMAIPOYLW 300  
 QY 820 TCTOEYVOSDYSSGPMRPFQHPKGFASGSPVD 854  
 DB 301 TCTOEYVOSDYSSGPMRPFQHPKGFASGSPVD 335  
 RESULT 8  
 ABP41342  
 ID ABP41342 standard; Protein: 249 AA.  
 XX  
 AC ABP41342;  
 XX  
 DT 23-AUG-2002 (first entry)  
 XX  
 DE Human ovarian antigen HFXLL31, SEQ ID NO:2474.  
 XX  
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antineoplastic; gynaecological; reproductive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200677-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 07-JUN-2001; 2001WO-US18569.  
 PR 07-JUN-2000; 2000US-209467P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Birse CE, Rosen CA;  
 DR WPI: 2002-147878/19.  
 DR N-PSDB; ABQ54419.  
 XX  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 XX  
 PS Claim 11; SEQ ID NO 2474; 2922pp; English.  
 XX  
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP41328) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and







Matches 80; Conservative 47; Mismatches 91; Indels 33; Gaps 7;

OY 57 GENYGMQIPNHNDF-----KILKN-----EROCEYLONKFGCISLVSFV 99  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

DB 359 GGGTTL--PHDGGGVKYNLYPEKLNEXYFOPKRVGFEELFERLRKLM-----PE 408  
 OY 100 OEGNSKSLQVFKMLTPRIELSVKMDLTTNADAVVNAANEDLHGGLALALVKGAGF 159  
 DB 409 KKGESMAVEKRELEPKKKIRIVGDIITREVDALVNAANXYLKHGGVACAIYRAGGS 468  
 OY 160 ELQESKQFVARYKVSAGELAVTGAGRLPKQIIHANGPRMMDKQCTCKLQRAIYS 219  
 DB 469 VLEESDRIVQERGRVPGGEAVTSAGKLAKAYVHTVGPMV-----RGSSEDELELYXA 524  
 OY 220 ILNVIYKYN-THIKTVAPALSSGIFPEPLNCTKTYIVTIVSLOGKPRMSLKEIHLY 278  
 DB 525 VYNALLRAHEMLKKSISMPAISTGIFGPKERAVGIFSKAINDFIDQHP-DITLLEIRIC 583  
 OY 279 SNEDPTVAAFK 289  
 DB 584 NIDEETTKIFE 594

RESULT 13  
 AAB94051  
 ID AAB94051 standard; Protein; 419 AA.

AC AAB94051;  
 DT 26-JUN-2001 (first entry)  
 DE Human protein sequence SEQ ID NO:14218.  
 KM Human: primer; detection; diagnosis; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 PN EPI074617-A2.  
 PD 07-FEB-2001.  
 PF 28-JUL-2000; 2000EP-0116126.  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 XX Claim 8; SEQ ID 14218; 2537bp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

DB 480 SSEMARRSKMLSLNNVSPQSTREKRENGLEAR--SPAINLMGFVDEMYFAHAMIORI 537  
 DB 10 SSQOSVMSKTLASFGLFS--KQSPQKKNHLYLEKTESATFRYCGENVTCVEAISMIDL 67  
 OY 538 LSL-QNHHTIENNHLIYGRKEHDILSOLQTSVSTETIISPGTELEIEGARADLIEV 596  
 DB 68 IEKQCPYTSDECEIKDPEDEKEYOELNELQKKLININIS--LDHKRPLIKVLGISRDVQA 125  
 OY 597 VMNIEDMLCKVQ-----EEMARKE--- 616  
 DB 126 RDEIEMIRKRVLAKEQESRADCSIEFTWQYNDNNTSHCFNKMTNLKLEARRRKKTV 185  
 OY 617 -----RGLMRSLGOWTIOQOKTODEMKNIEFLKCPVP-----TQELLDQK 659  
 DB 186 DVKINRHRYTVALNLTNTADTGHSLSVQRLKSVYDIPAHMSDKQONFCVELLPSP 245  
 OY 660 QF-----EKCG-LQYLKVEKIDNEVLMAAFQKKKMEKLLHROPYSRFLPQ----- 705  
 DB 246 EYNTVASKFNQTCSHREIKERIQNPDLMSYQAKKTMADKNQJNEKOLFPHGTAG 305  
 OY 706 QVPYQECNVYCVGFQPMYSPCDPRKYGAGIYFTKLNKLAKAKKISADKLIVFEAE 765  
 DB 306 SVPH-----VNRNGFRKRSIAGKNAVAAYGKGTFFAANANYSANDYTSRPNARKKHYYR 360  
 OY 766 VLTGFEQGHPIINIVPPLSP-GAIDGHSVDVNVSSPETFAIFSGMAIPIQYLMT 820  
 DB 361 VLTGIYTHGNHSLIVPPSKNPQNPDLVDTVDVNHHPRLFAFYDYQAYPEYLIT 416

RESULT 14  
 ABB49353  
 ID ABB49353 standard; Protein; 176 AA.

AC ABB49353;  
 DT 05-FEB-2002 (first entry)  
 DE Listeria monocytogenes protein #2057.  
 KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease.  
 OS Listeria monocytogenes.  
 PN W0200177335-A2.  
 PD 18-OCT-2001.  
 PF 11-APR-2001; 2001WO-FR01118.  
 PR 11-APR-2000; 2000FR-0004629.  
 PA (INSP ) INST PASTEUR.

PI Buchrieser C, Frangoul L, Couve E, Rusnick C, Esli H, Dehoux P,  
 PI Dussurget O, Cheonani F, Nedjati H, Glaser P, Kunst F, Cossart P,  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
 PI Chakraborty T, Domm E, Hain T, Berche P, Charbit A, Durant L,  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-lopez N,  
 PI Madueno E, de Pablos B, Wehland J, Kaerst U, Entian K, Hauf J,  
 PI Rose M, Voss H;  
 WPI: 2002-010914/01.  
 DR  
 XX  
 XX  
 PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides -  
 PS  
 XX  
 PS Claim 6; SEQ ID NO 2058; 192pp; French.  
 CC  
 XX  
 CC The present invention relates to the genome sequence of *Listeria*  
 CC *monocytogenes* ESD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC *monocytogenes* and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of *L. monocytogenes* and related organisms,  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by *L.*  
 CC *monocytogenes* and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX  
 XX Sequence 176 AA;  
 SO  
 Query Match 5.3%; Score 235; DB 23; Length 176;  
 Best Local Similarity 32.2%; Pred. No. 1.1e-11;  
 Matches 55; Conservative 39; Mismatches 73; Indels 4; Gaps 3;  
 OY 118 IELSYWKDLDLTHAVDAVANNANEDLHGGLALALVRAGFEIOESKOFVARYKSYA 177  
 DB 1 MEITIVKGDITPEQEVDAIVNANNSGLLGCGVDGAIHQAGADLLKEQOEIVNRIGSCPA 60  
 OY 178 GEIANTGAGRLPCKOITIAVGEPRMWMKQCGCTKLRRAIVSILNVIYKNTHTKTAIP 237  
 DB 61 GEAVTISAGDLKATYIIHAVGPIMKDGHOE-ANKLASCYWKALDLAAGKD--LTSIAFP 117  
 OY 238 ALLSGIFQFPLNLCTKTIVETIRVSLQCKPMNSNKEIHLVSNEDPTVAAP 288  
 DB 118 NISTGVGPKKLAVALVYVRKMAE-EYDTSIKETIRFVCEDEENLKY 167  
 RESULT 15  
 ID AAU58306  
 ID AAU58306 standard; Protein: 246 AA.  
 XX  
 AC AAU58306;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #19202.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopthic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN W0200181581-A2.

XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIAX CORP.  
 XX  
 PI Skeiky YAN, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L.maisonueve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI: 2001-616774/71.  
 DR N-PSDB: AAS59590.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 PS  
 XX  
 PS Example 1; SEQ ID NO 19501; 1069pp; English.  
 CC  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC *P. acnes* is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of *P. acnes* in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for *P. acnes* proteins. These antibodies can be used to  
 CC downregulate expression and activity of *P. acnes* polypeptides and  
 CC therefore treat *P. acnes* infections. The antibodies may also be used as  
 CC diagnostic agents for determining *P. acnes* presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX  
 XX Sequence 246 AA;  
 SO  
 Query Match 4.9%; Score 219; DB 22; Length 246;  
 Best Local Similarity 27.2%; Pred. No. 4.1e-10;  
 Matches 61; Conservative 46; Mismatches 89; Indels 28; Gaps 6;  
 OY 84 VLONKFCISTSLVSEVQGNKSLQVFRKMLTPRIELSVWKDLDLTHAVDAVANNANEDL 143  
 DB 48 LLNARSASIAVTPDPSAEGBGSCSTI-----GAMADITLLRADITLTDVDAVANNANQOL 102  
 OY 144 LHGGGLALALVRAGFEIOESKOF--VARYKVSAGIANTGAGRLPCKOITIAVGEPRW 201  
 DB 103 AGGCGVDGAIHQAGADLLKEQOEIVNRIGSCPA 60  
 OY 202 MEW-DKQCGCTKLRRAIVSILNVIYKNTHTKTAIPALLSGIFQFPLNLCTKTIVETIR 260  
 DB 163 AKTIDKSDQLASCYRTSLHVADEI-----GARTIAFPITSAGVGYGPMDEATRLAVENC 217  
 OY 261 VSLQCKPMNSNKEIHLVSNEDPTVAAPKASFTILCKSELGQF 304  
 DB 218 QT-----VTKVDITLYLVAFAEAGYRAA-----LGEE 246  
 Search completed: May 13, 2003, 11:42:12  
 Job time : 86 secs

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:40:47 ; Search time 30 Seconds  
(without alignments)  
837.573 Million cell updates/sec

Title: US-09-830-762-2

Perfect score: 4447  
Sequence: 1 MDESNVAGAAAYNEKSGRIT.....PMRPQHPHWRGFSAGSPVD 854

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCFUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 165.5 | 3.7         | 368    | 2     | US-08-824-878-3    |
| 2          | 165.5 | 3.7         | 368    | 4     | US-09-353-688-3    |
| 3          | 159   | 3.6         | 373    | 2     | US-08-824-878-1    |
| 4          | 159   | 3.6         | 373    | 4     | US-09-353-688-1    |
| 5          | 136.5 | 3.1         | 1375   | 4     | US-09-722-139-2    |
| 6          | 136.5 | 3.1         | 1375   | 4     | US-09-721-832-2    |
| 7          | 136.5 | 3.1         | 1375   | 4     | US-09-721-689-2    |
| 8          | 134   | 3.0         | 1166   | 4     | US-09-350-982C-5   |
| 9          | 133   | 3.0         | 1786   | 4     | US-08-973-462-8    |
| 10         | 132.5 | 3.0         | 2512   | 2     | US-08-801-263A-9   |
| 11         | 132.5 | 3.0         | 2512   | 4     | US-09-102-248-9    |
| 12         | 132   | 2.9         | 1327   | 4     | US-09-196-387-2    |
| 13         | 131   | 2.9         | 2517   | 2     | US-08-801-263A-5   |
| 14         | 131   | 2.9         | 2517   | 3     | US-09-102-248-5    |
| 15         | 130.5 | 2.9         | 712    | 2     | US-08-468-576B-17  |
| 16         | 130.5 | 2.9         | 712    | 3     | US-08-468-577B-17  |
| 17         | 130.5 | 2.9         | 712    | 3     | US-08-468-577B-17  |
| 18         | 130.5 | 2.9         | 2500   | 2     | US-08-801-263A-2   |
| 19         | 130.5 | 2.9         | 2500   | 3     | US-09-102-248-2    |
| 20         | 118   | 2.7         | 498    | 4     | US-08-858-207A-310 |
| 21         | 118   | 2.7         | 2860   | 2     | US-08-826-267-2    |
| 22         | 116   | 2.6         | 2037   | 4     | US-09-306-998-3    |
| 23         | 115.5 | 2.6         | 630    | 4     | US-08-973-462-9    |
| 24         | 114.5 | 2.6         | 1013   | 4     | US-09-255-829-18   |
| 25         | 114   | 2.6         | 397    | 6     | 5187089-2          |
| 26         | 114   | 2.6         | 397    | 6     | 5187089-9          |
| 27         | 114   | 2.6         | 397    | 6     | 5187089-10         |

|    |       |     |      |   |                     |                    |
|----|-------|-----|------|---|---------------------|--------------------|
| 28 | 114   | 2.6 | 397  | 6 | 5187089-11          | Patent No. 5187089 |
| 29 | 114   | 2.6 | 1053 | 4 | US-09-724-519-2     | Sequence 2, App1   |
| 30 | 114   | 2.6 | 1053 | 4 | US-09-592-037-2     | Sequence 2, App1   |
| 31 | 112   | 2.5 | 1038 | 4 | US-09-541-782-4     | Sequence 4, App1   |
| 32 | 112   | 2.5 | 1038 | 4 | US-09-723-820-4     | Sequence 4, App1   |
| 33 | 112   | 2.5 | 1388 | 4 | US-09-572-191-2     | Sequence 2, App1   |
| 34 | 112   | 2.5 | 1388 | 4 | US-09-723-262-2     | Sequence 2, App1   |
| 35 | 112   | 2.5 | 1388 | 4 | US-09-723-219-2     | Sequence 2, App1   |
| 36 | 111.5 | 2.5 | 514  | 2 | US-08-960-023-14    | Sequence 14, App1  |
| 37 | 111.5 | 2.5 | 550  | 2 | US-08-844-058-2     | Sequence 2, App1   |
| 38 | 111.5 | 2.5 | 3696 | 4 | US-09-134-001C-5080 | Sequence 5080, Ap  |
| 39 | 111   | 2.5 | 397  | 6 | 5457090-2           | Patent No. 5457090 |
| 40 | 111   | 2.5 | 397  | 6 | 5457090-4           | Patent No. 5457090 |
| 41 | 111   | 2.5 | 397  | 6 | 5495001-7           | Patent No. 5495001 |
| 42 | 110.5 | 2.5 | 1589 | 3 | US-08-755-587-189   | Sequence 189, App  |
| 43 | 110   | 2.5 | 3248 | 1 | US-08-353-700-1     | Sequence 1, App1   |
| 44 | 110   | 2.5 | 3248 | 5 | PCT-US85-16216-1    | Sequence 1, App1   |
| 45 | 109.5 | 2.5 | 398  | 6 | 5187089-4           | Patent No. 5187089 |

#### ALIGNMENTS

RESULT 1  
US-08-824-878-3  
; Sequence 3, Application US/08824878  
; Patent No. 5981221  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: NOVEL HISTONE FUSION PROT  
; TITLE OF INVENTION: E1N  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/824,878  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0255 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 368 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: Genbank  
; CLONE: 205276  
; US-08-824-878-3

Query Match 3.7% Score 165.5; DB 2; Length 368;  
Best Local Similarity 24.1% Pred. No. 8e-07;  
Matches 76; Conservative 47; Mismatches 140; Indels 53; Gaps 10;

```

Oy      6 VAGAAVAENKSGRITSLSLLPOKFAOIFPMRGNT-----ECLPKSGSETG 54
Db      63 LAGNAARONKKGWYPRILLAVANDEBNOLKGVITASGVLPNTNHPLELKKRSKG 1222
Oy      55 AL-----GENSMQIPRINHDEKLKNEROLCEVLONKFCISITLVS--PYOE 1010
Db      123 KLEAITPPPAKKKASPSQKKPVAKKTGG--KKGARK----SKQOEVSKASASBSTE 175
Oy      102 G-----NSKSLOVRRKMILPRIELTSYWKDDLTTHAVDAVAVNANDELLHGGGALA 152
Db      176 GAPTGGFTYLTSTKSLFLGQK-----LQVQADIAISDSAAVHNPTNDFYIGGEVST 228
Oy      153 LKAGGFTQESKQOFVARIVKVSANGELVAITGAGRLPCKOIIHAVGRMMEBMDKQGTGK 212
Db      229 LEKKGKGFEEVAUVELLRKKNRPLEVAGAAVASAGHLPAKFIHONSP--VWSDKCEL 285
Oy      213 LORALVSLINLVIAKNTIKITVAIPALSSGIFOPMLNICTTIVETIRVSLQCKPMKSNL 272
Db      286 LEKTVKNCNL--ALDDRKLRKSIARPSIGSGNGNRPKQTAQMLIKALIS--SYFVSTMSSEI 342
Oy      273 KEIHLYSNEDPTVAAF 288
Db      343 KTVYFVLEFDESIGLY 358

```

RESULT 2  
US-09-353-688-3  
: Sequence 3, Application US/09353688

GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: NOVEL HISTONE FUSION PROTEIN  
TITLE OF INVENTION: EFN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/353,688  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/824,878  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0255 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 368 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 205276  
US-09-353-688-3

|                       |        |                  |       |             |
|-----------------------|--------|------------------|-------|-------------|
| Query Match           | 3.78;  | Score 165.5;     | DB 4; | length 368; |
| Best Local Similarity | 24.18; | Pred. NO. 8e-07; |       |             |

| Matches | 76; Conservative  | 47; Mismatches | 140; Indels | 53; Gaps | 10; Nucleotide |
|---------|---|----------------|-------------|----------|----------------|
| QY      | 6 VAGAAAYNEKSGRITSLLEFQVFAQIFPQRRKNT-----ECLPYKCSQETG               | 54             |             |          |                |
| DB      | 63 LAGNAARQDKKGRVPRHILLAVANDEELNQLKGVITIASGCVLPNIHPELLAKRKSGK       | 122            |             |          |                |
| QY      | 55 AL-----GENYSNQIPININDPFIILKNNEQOLCEVQNKFGCSTLVS--PYQE            | 101            |             |          |                |
| DB      | 123 KKEAIIIPPAKKAASPSQKPPARKTG--KKGAR-----SKQGEVSKAASADSTTE         | 175            |             |          |                |
| QY      | 102 G-----NSKSLQVFQKMLTPRIETLSVWKDDITTHAVDAVVAANAANDLLHGGLALA       | 152            |             |          |                |
| DB      | 176 GAPTQGFYTLSTKSLFLGQ-----LQVQADIASIDSDAVHPTNTDFYIGGEVSGT         | 228            |             |          |                |
| QY      | 153 LVKAGGFELQEESSQFVARVGRVSGAGELAVTGAGRPCQKQIITHAVGPRMWMEDKQGGCTGK | 212            |             |          |                |
| DB      | 229 LEKKGSGKEFVEAVLELRKKNGPLEVAGAAVSGHGIAPAFVHCNSP--VWGSQKCEEL      | 285            |             |          |                |
| QY      | 213 LQRAVSLINVIYKNTHTKVTAFALSSGIRQFPLNLTCTKIVETIRISLQGGKPRMSML      | 272            |             |          |                |
| DB      | 286 LEKTVKNCI--ALADRRKLKSLAFPSIGSGRRGFPKQTAADILKLAIS--SYFVSTSSSSI   | 342            |             |          |                |
| QY      | 273 KEIHLVSNDDPTVAAF  | 288            |             |          |                |
| DB      | 343 KTVFVLFQDSISIGIY  | 358            |             |          |                |

### RESULT 3

; Sequence 1, Application US/08824878

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

US-08-824-878-1



```

05-09-353-688-1
Query Match. 3.6%; Score 159; DB 4; Length 373;
Best Local Similarity 22.7%; Pred. No. 3,3e-06;
Matches 72; Conservative 54; Mismatches 141; Indels 50; Gaps 9.

QY 6 VAGAAVNEKSGRITSLLEFQVFAQIIFPQWRKNT-----EECLPYKCEHG 54
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 63 LAGNAARONKKGGRVPRPHILLAVANDEDELNOLLKGVTTIASGGVLRPIHPELLAKKRGSG 122
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 55 AL-----GENYSQIPIINHDEKILKNRROLCEVLNRCGICSTLVS----- 98
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 123 KLEAITTPPAKKAKASPSQKKRFSK-----KAGKKGKARKKSKKKGVEKSKAASDSNR 176
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 99 ---VOEG-----NSKSLQVFRKMLTPRIELISVWKDLTTTHAVDAVYNAANEDLLHGGLAL 151
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 177 GEHLPDGFTVLSTKSLFLGQKLNLIHSEIS-----NLAGEVEEALINPTVADIDLKDDLGN 232
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 152 ALVKAAGEFIEODESKQFVARYGKVSAGELIAYTGAGRLPCQIITHAVGPRMMEWDKOGCTG 211
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 233 TLEKKGKKEFEVAVLELKRKNKNGLEVAGAAVSAGHGILPAKFVFIHCNSP---VWGDADKCE 289
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 212 KLGRAVSLINLVITIKKNHIKTVALPALLSSGIFQPPNLCTIYETTVSLQGRPMMSN 271
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 290 PLEKTVKNCCL--ALADDKKLKSIAEPSIGSGNGFPKQTAOLILKALS--SYFVSMTSSS 346
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 272 LKEIHLVSNEDPTVAAF 288
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 347 IKTIVFVLDPDESIGIY 363
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

```

RESULT 5
US-09-722-139-2
; Sequence 2, Application US/09722139
; Patent No. 6355471
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6355471el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1055
; CURRENT APPLICATION NUMBER: US/09/722.139
; CURRENT FILING DATE: 2000-11-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Human
US-09-722-139-2

Query Match      3.1%; Score 136.5; DB 4; Length 1375;
Best Local Similarity 18.7%; Pred. No. 0.0033;
Matches 154; Conservative 131; Mismatches 299; Indels 239; Gaps 38

QY      2 DFSWVAGAAVNEKSGRTSLSLFLQKVFQAQTFPMWRKGNTEECLEPYKCSFTGALGENYS 61
      111      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59 DSEFYSA---DTKSPDYVSOEMVFKTLGTDVKSAFEGYNACVFAY-----GQTGSGKS 109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      62 WOIPINHDFKLKNNEROLCEVLONKF-----GCSTLSPVOEGNSKSLQYFRKM 113
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      110 YTMGNSGSGSLIP---RICEGLFSRINETTRMWDASRFTVSTLEYINERVDDLRRK 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      114 LTRPRLSV-----WKDDLTTTHAVDAVVAANEDLLHGGGLALALVKAGGFETJOES 165
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      166 SSKTFNLRLVREHPKEGYPVEDLSKILYONYGDV--EELMDAGNINRTTAATGMNDVSSRS 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      166 KQFVARIKVSGELFANVGA---GMLPKQI--THAVGRPMNE-WDKQGGT-----CK 212
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      224 HAIFV-----IKFTQAKFDSMPCEYKSIHLVLDAGSERADATGATVRLKEGNN 274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      213 LQRAIV-----SLNLYVYKNT--HIKTVAIPALSSGIFQFPLNLT--- 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 275 INKSLVLTGAKKQVFPYRDSVLTWLLKDSLGSKTIIMATISPADVNGETLSLRY 334  
QY 233 -----KTIYETIRVSL--OGKPMNSLKEIHLVSN-----EDPTVAFAKASEFI 295  
Db 335 ANRAKNIINKPTINEDANVKLIRELRAIARLKTLLAOGNIALDSTALSMEEKLQON 394  
QY 296 LGR-SELGOETTPSFNAN--VYNNLTLOIYOGHIEMOTADVIYNSVNHDTIVGPVAKSI 352  
Db 395 EARVOELTKEMTNKMNENTONILKEOTLALRREGI-----GVLDSELPHLI----- 440  
QY 353 LQOAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVLMHSEF-----PKPOI 405  
Db 441 ----GID--DDLSTGI-----ILYHLKEGQTYVGRDASTEDDI 474  
QY 406 LKHAMKECLEKICIONT--TSISFPALGT-----GNMEIKKET-----AAELFDEVLPFAK 455  
Db 475 VLHGIDLESEHCIFENIGCTVTLPLSGSQCYSVNGVOIYEATHLNOGAVILLGRTNMF 534  
QY 456 DHVKHQLTVKRVFIPPTDEIYKAFSESEMAKRSKMLSNNSVPOSTREKREN----- 508  
Db 535 NHPKE-----AAKLRKRKSGLLSSFSLSMTDLKSRENLSAVMLY 575  
QY 509 --GLE-ARSPAINLMGFN-----VEEMEYEAH--AWIOTRI-----LSLQ 541  
Db 576 NPGLEFEHQOELELEKLESKRKLIEEMEKOKSDKAELERMOQEVEYORKETEIYOLQIR 635  
QY 542 NHHIENNNHILYLRKREDILSOLQTSVSITETIISPGTELEEGARADILEVANNIE 601  
Db 636 KOESLKRSPHIEKLDLAELKEKFEERLRE-----QOEILOKKRQF-----E 682  
QY 602 DMLCKVOEMARKKRGILMSLQGTI-----OOQKODEKKNENILFKCVPPTQELLD 656  
Db 683 ETLFVQOELORLKELNNEKAERKFOIFQELDOLEKEDEQYAKLELEK-----KRL 736  
QY 657 OKKOFKCGLOVLVYKIDNEVLMAAFQRRKKMMEEKLHROPV 699  
Db 737 OEKE-----QVMLVAHLEQL-----REKQEMIQLLRGEV 767

RESULT 6  
US-09-721-832-2  
; Sequence 2, Application US/09721832  
; Patent No. 6399346  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: No. 6399346e1 motor proteins and methods for  
; FILE REFERENCE: 1055  
; CURRENT APPLICATION NUMBER: US/09/721,832  
; CURRENT FILING DATE: 2000-11-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Human  
US-09-721-832-2

Query Match 3.1%; Score 136.5; DB 4; Length 1375;  
Best Local Similarity 18.7%; Pred. No. 0.0033;  
Matches 154; Conservative 131; Mismatches 299; Indels 239; Gaps 38;

QY 2 DFSWAGAAVNEKSGRTISLSLFQKVFQIIFQWRKGNTECLPYKCSFTGALGENYS 61  
Db 59 DFSFYSA---DTKSPDYVSOEWKFTLGTIDVYKSAFEGYACVFAY-----GOTGSGKS 109  
QY 62 WQIPININDFKILKNNEPOLCEVLQNKF-----GCISLTVSPVQEGNSKSLQVFRKM 113  
Db 110 YTMNGNSGDSQLP-----RDEGLFSKRNTETRWDEASFTREVSLEYLYENKRVLDLRRK 165  
QY 114 LTPRIELSV-----WKDULTTHAVDAVVAANANEDLLHGGGLALVLVKGGEFIOEES 165

Db 166 SSKTFNLVRREHPKPGPYEDLSKHLVONYGDV--EELMDAGNINRTTAATGMDVSSRS 223  
QY 166 KOFVARYGKVASGEIAYVGA---GRLPCKQI--IHAVGPRWME--WDKOGCT-----GK 212  
Db 224 HAIF-----IKFTQAFKFDSEMPCEYVSKHVLVLASSERDATTGATGVRLLKEGN 274  
QY 213 LQRAIV-----SILNVIYIKNT--HIKTVAILPSSGIFQFPLNCT--- 252  
Db 275 INKSLVLTGAKKQVFPYRDSVLTWLLKDSLGSKTIIMATISPADVNGETLSLRY 334  
QY 253 -----KTIYETIRVSL--OGKPMNSLKEIHLVSN-----EDPTVAFAKASEFI 295  
Db 335 ANRAKNIINKPTINEDANVKLIRELRAIARLKTLLAOGNIALDSTALSMEEKLQON 394  
QY 296 LGR-SELGOETTPSFNAN--VYNNLTLOIYOGHIEMOTADVIYNSVNHDTIVGPVAKSI 352  
Db 395 EARVOELTKEMTNKMNENTONILKEOTLALRREGI-----GVLDSELPHLI----- 440  
QY 353 LQOAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVLMHSEF-----PKPOI 405  
Db 441 ----GID--DDLSTGI-----ILYHLKEGQTYVGRDASTEDDI 474  
QY 406 LKHAMKECLEKICIONT--TSISFPALGT-----GNMEIKKET-----AAELFDEVLPFAK 455  
Db 475 VLHGIDLESEHCIFENIGCTVTLPLSGSQCYSVNGVOIYEATHLNOGAVILLGRTNMF 534  
QY 456 DHVKHQLTVKRVFIPPTDEIYKAFSESEMAKRSKMLSNNSVPOSTREKREN----- 508  
Db 535 NHPKE-----AAKLRKRKSGLLSSFSLSMTDLKSRENLSAVMLY 575  
QY 509 --GLE-ARSPAINLMGFN-----VEEMEYEAH--AWIOTRI-----LSLQ 541  
Db 576 NPGLEFEHQOELELEKLESKRKLIEEMEKOKSDKAELERMOQEVEYORKETEIYOLQIR 635  
QY 542 NHHIENNNHILYLRKREDILSOLQTSVSITETIISPGTELEEGARADILEVANNIE 601  
Db 636 KOESLKRSPHIEKLDLAELKEKFEERLRE-----QOEILOKKRQF-----E 682  
QY 602 DMLCKVOEMARKKRGILMSLQGTI-----OOQKODEKKNENILFKCVPPTQELLD 656  
Db 683 ETLFVQOELORLKELNNEKAERKFOIFQELDOLEKEDEQYAKLELEK-----KRL 736  
QY 657 OKKOFKCGLOVLVYKIDNEVLMAAFQRRKKMMEEKLHROPV 699  
Db 737 OEKE-----QVMLVAHLEQL-----REKQEMIQLLRGEV 767

RESULT 7  
US-09-721-689-2  
; Sequence 2, Application US/09721689  
; Patent No. 6440685  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: No. 6440685e1 motor proteins and methods for  
; FILE REFERENCE: 1055  
; CURRENT APPLICATION NUMBER: US/09/721,689  
; CURRENT FILING DATE: 2000-11-24  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Human  
US-09-721-689-2

Query Match 3.1%; Score 136.5; DB 4; Length 1375;  
Best Local Similarity 18.7%; Pred. No. 0.0033;  
Matches 154; Conservative 131; Mismatches 299; Indels 239; Gaps 38;

QY 2 DFSWAGAAVNEKSGRTISLSLFQKVFQIIFQWRKGNTECLPYKCSFTGALGENYS 61  
Db 59 DFSFYSA---DTKSPDYVSOEWKFTLGTIDVYKSAFEGYACVFAY-----GOTGSGKS 109  
QY 62 WQIPININDFKILKNNEPOLCEVLQNKF-----GCISLTVSPVQEGNSKSLQVFRKM 113  
Db 110 YTMNGNSGDSQLP-----RDEGLFSKRNTETRWDEASFTREVSLEYLYENKRVLDLRRK 165  
QY 114 LTPRIELSV-----WKDULTTHAVDAVVAANANEDLLHGGGLALVLVKGGEFIOEES 165

```

Db 59 DSEFYSA----DTRSPDYVSOEMVFKLTGTDVYKSAFEGYNACVFAY-----GOTSGSKS 109
QY 62 WOIPINHDEFLKKNNEKROCEVLONKF-----GCISTLVSPVCGNSKSLQVFKRM 113
Db 110 YTMGNNSDGLIP-----RICEGLFSRINETTRMDEASFRIEVSLEYLNERVRLRK 165
QY 114 LTPRIELSV-----WKDDLTHAVDAVYVNAANEDLLHGGGLALVLVKAQGEFIOEBS 165
Db 166 SSKTFENLREHPKRGPEYVEDLSKLYQNGDV--EELMDAGNINRTTAAGMNDVSSRS 223
QY 166 KQFARIVKVSAGELAVTGA--GRLPCKQI--IHAVPRRME-WDKQCGT-----GK 212
Db 224 HAIFP-----IKFTQAKFDSSEMPCTVSKHLVDLASERDADATGATGRLEKGN 274
QY 213 LQRAIV-----SILNVYLYKMT--HIKTVAPALSSGIFQPLNCT-- 252
Db 275 INKSLVITGAKKQVYFVYRDSVLTWLLKDSLGNSKSTIMATISPADVNGETILSTLRV 334
QY 253 -----KTIYETIRVSL--QCKPMMSNLKEIHVSN-----EDPTVAFAKASEFI 295
Db 335 ANRAKNIINKPLINEDAVVKLIRELRAIARLKTLLAGNOALDPSPTALSMEEKLQON 394
QY 296 LQK-SELGOETPSFNAM--VYNNLTLOIYOGHIEMOTADYIVNSVNHDTVGVVAKSI 352
Db 395 EAVVOELTKENTNKNWNETONILKEOTLARKEGI--GVVLDSELPHLI----- 440
QY 353 LQAGVEMKSEFLATKAKOFORSOLVLTGKFNLECKYIVHLMSEF-----PKPOI 405
Db 441 ----GID--DDLSTGI-----ILYHKEGQTVVGHDDASTBEDI 474
QY 406 LKHAKECELEKICIONT--TSTISPALGT-----GMELTKET--AAELIFDEVLTAK 455
Db 475 VLHGIDLESEHCIFENIGITVTLPLDSGOSCVNGOIVEATHLNOGAVILLGRTNMRP 534
QY 456 DHVKHQLVFKVIFPTDLEIKAPSEEMAKSKMLSNNSYPOSTREKREN----- 508
Db 535 NHPRE-----AAKREKRKSGLLSSFSLSMTDLSKSRNLSAVMLY 575
QY 509 --GLE-ARSPAINLMGFN-----VEEMVDAH--AWTORI-----LSIQ 541
Db 576 NPGLEFERQORELEKLSKRKRIEMEMEKOKSKAKELERMOQOEYETQKREIYQDLIR 635
QY 542 NHHIENNHHILYKKEKEDIISQLOKTSVSTTEIISPGRIELEEGARADILEVYVNI 601
Db 636 KOESILKRRSRHIEKMLDLAEKKEFEEERLRE-----QOEIELOKKROE-----E 682
QY 602 DMLCVQOEMARKKRGMLRSLSGWTI-----QOQKTODEKMENTIIFLKCVPPTQELLD 656
Db 663 EFTLVQOELQKELNNNEKAKEKFOIPELDQLOKEKEDQYAKLELEK-----KRLKE 736
QY 657 OKKOFKCGLOVLKVEKIDNEVYLAFAFORKKKMEKELHROPV 699
Db 737 QEKE-----QYMLVANHLEBOL-----REKQEMTOLLRRGEV 767

```

```

RESULT 8
US-09-350-982C-5
: Sequence 5, Application US/09350982C
: Patent No. 6455290
: GENERAL INFORMATION:
: APPLICANT: Bertheisen, Jens
: APPLICANT: Toma, Salvatore
: APPLICANT: Isaacchi, Antonella
: TITLE OF INVENTION: Tankyrase Homolog Protein(THP), Nucleic Acids, and Methods Rel
: TITLE OF INVENTION: Same
: FILE REFERENCE: PHM-0043
: CURRENT APPLICATION NUMBER: US/09/350, 982C
: CURRENT FILING DATE: 1999-07-09
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 5
: LENGTH: 1166
: TYPE: PRT

```

```

: ORGANISM: Artificial
: FEATURE:
: NAME/KEY: misc:feature
: OTHER INFORMATION: Xaa is any amino acid
: NAME/KEY: misc:feature
: LOCATION: (1102)..(1102)
: OTHER INFORMATION: n is any nucleic acid
: NAME/KEY: misc:feature
: LOCATION: (2650)..(2650)
: OTHER INFORMATION: n is any nucleic acid
US-09-350-982C-5

```

```

Query Match
Best Local Similarity 3.0%; Score 134; DB 4; Length 1166;
Matches 43; Conservative 28; Mismatches 74; Indels 24; Gaps 6;

```

```

QY 668 VLKVEKIDNEVYLAFAFORKKKMEKELHROPVSHRLEFOOVYQPCNVYCRVGFQRMSTP 727
Db 994 ILKIQVCNKKLMERYTHRRKREVESENNHANERMLFPGSP--FVNALIRKGFERNAY- 1050
QY 728 CDPKYGAGIYFKNLKNLAEKAKKTSAA-----DKLIYFEAEVLGFPK----- 772
Db 1051 ICGMFGAGIYFRAENSSKSNQYVYIGGTCGCPVHKDRSCYICHRQL--PCRYTLKXSF 1107
QY 773 -QGAPLNIYPPPLSPGALDGHDSVDVNVSSPTEFYFSGMOAIPQYLMT 820
Db 1108 LQFSAMKNAHSPGHHSVTGRPSV--NGLALAEYVYRGEQAVPEYLLT 1154

```

```

RESULT 9
US-08-973-462-8
: Sequence 8, Application US/08973462B
: Patent No. 6191270
: GENERAL INFORMATION:
: APPLICANT: DRUTHE, PIERRE
: APPLICANT: DAUBERSTES, PIERRE
: TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
: FILE REFERENCE: 0660-0125-0 PCT
: CURRENT APPLICATION NUMBER: US/08/973, 462B
: CURRENT FILING DATE: 1998-02-06
: EARLIER APPLICATION NUMBER: PCT/FR96/00894
: EARLIER FILING DATE: 1996-06-12
: EARLIER APPLICATION NUMBER: FR 95/07007
: EARLIER FILING DATE: 1995-06-13
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 1786
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

```

Query Match
Best Local Similarity 3.0%; Score 133; DB 4; Length 1786;
Matches 124; Conservative 128; Mismatches 264; Indels 156; Gaps 28;

```

```

QY 81 LCEVLONKFCSTLVSPVOEGNSKSLQYFRKMLPRIELSWKDDLTTHAVDAVYVNAAN 140
Db 845 LNEIEVKEVNVVTTLENEVETTAESVTFSSNLE-----EIQENTTINDTIEEKL----- 895
QY 141 EDLLHGGGLALALVAGGFEIOEESKQFARVKGVSAGEIAVTGARGRLPKOIIHAVGPR 200
Db 896 -ELHENVLSALEMT--QSEEEKVEYDIVEYKE--EVATT-----LIETV--- 938
QY 201 WMEWDKQGTGKLQRAIVSILNVYLYKNTHTIKTVAIIPALSSGIFQPLNCTKTIYETIR 260
Db 939 --EQAEKKSANTITEIFEENLAEVASENENV--AENLEKINTEVETVLDKVEET--VEISG 994
QY 261 VSLQCKPM-----MSLTKRIHIVSNEDPPIVAAFKAASEIILKSKSELGOETTPSFNAM 312
Db 995 ESLENNEMDKAFSEIFDNVKGI---QENLLTGMFRSITETSIYQSEKQVLDNENYSS 1050

```

```

QY 313 VVNNLTQIVGHIEMOTADVIIVNSVPHDITVG---PVAKSILQOAGVE-----MKSEF 364
DB 1051 ILDNIE-----NMKEGLINKLENISSTEGVOETVENOVYVDVVPAMKDOF 1100
QY 365 LATKAKOFORSOLVLTGK-----FNL--FCKYIYHVLHNSFPKPOLIKHAMKECLF 415
DB 1101 LG-----ILNEAGGLKEMEFNLEDVFKSESVDYTVBEIKDEPVQKEVEKETVSI 1149
QY 416 -KCIBONITISIFPALGNGNMEIKKETAEILFD---EVLTRAKDHVKHOLTIVKFIPT 471
DB 1150 IEMENIVDV-----LEEKEDLTDMKIDAVEESIEISSKETEESIKKEKEDV 1200
QY 472 DLEIKAFSESSEMAKR-SKMLSLNNYSVPOSTREKRENGLEARSPAINLMGFVEMWEA 530
DB 1201 SLIVEEVDNDMDSEVKEVLKKN-----MEELMKDAVELINDITSKIE-ETDELNV 1253
QY 531 HA-----WIORIISLONHNIIE--NNHILYLRKREHDITLSOLKTSSVSTIEIIS 578
DB 1254 EADLIKMEKLEKLEKALSESEKEIIDAKDITLEKVEIEEHDTITLDEV----- 1303
QY 579 PGRTELEGARADLIEVVMNEDMLCKVOEEMARKKERGLMRSLGQWTIOQOKTQDEMK 638
DB 1304 -----VELKDVEEDKIEKVYSDLKLEEDILKEVKEIKE-----LESEIL 1342
QY 639 ENIFLKCPVPPTQELDOKKOFKCGLOVLKVE-----KIDNEVLMMAFQRRKKM 689
DB 1343 EDYKELK---TETDILKEKKEIEKDHFEKEFEAEIKDLLEADILKEVSSLEVEEKKL 1399
QY 690 MEKLRHROPVSH 701
DB 1400 EEVHELKEEVER 1411

```

# RESULT 10

US-08-801-263A-9

Sequence 9, Application US/08801263A

Patent No. 5811407

GENERAL INFORMATION:

APPLICANT: Johnston, Robert E.

APPLICANT: Davis, Nancy L.

APPLICANT: Simpson, Dennis A.

TITLE OF INVENTION: System for the In Vivo Delivery and

NUMBER OF SEQUENCES: 12 Expression of Heterologous Genes in the Bone Marrow

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bell Seltzer Park & Gibson, P.A.

STREET: 1211 East Morehead Street

CITY: Charlotte

STATE: No. 3811407th Carolina

COUNTRY: USA

ZIP: 28234

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/801,263A

FILING DATE: 19-FEB-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Sibley, Kenneth D.

REGISTRATION NUMBER: 31,665

REFERENCE/DOCKET NUMBER: 5470-147

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 2512 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-801-263A-9

Query Match

Best Local Similarity 3.0%; Score 132.5; DB 2; Length 2512;

Matches 72; Conservative 45; Mismatches 107; Indels 89; Gaps 13;

```

QY 26 FCKVAFQIFPCRRKNGTECELPY-----KCEFGALGENTSMQIPINHDFKILKN----- 76
DB 1227 YDLVFNIGTGRKRNHNFQCCEDHAATLEKTLRSALN-----CLNPGCTLVVKSGYAD 1279
QY 77 -NEROLCEVLQKFGCISTVLVSPVOEGNSKSLQVPRKM-----LTDR---ILSYW- 123
DB 1280 RNSEVDYVALAKFVRYGASARDCVSSNTEMLIFRDLNSTRQTPRTHNLNCVSSYVE 1339
QY 124 -----KDLITHAVDVYVNAANEDILHGGGLALVYKAGFEIOESKOF 168
DB 1340 GTRDVGGAAPSYRTKRENINADQEEAVVNAANPLGRPEGVCRAIYK----- 1386
QY 169 VARYGKVGAGELAVYAGARLP-C--KQIIHAVGPRMMDKQCGTGLQRAIVSLNTYVI 225
DB 1387 -RWPTSTTDSATETGTARMVYCLGKRVIIHVGPDPRKHPAEALKLQNAVHVAADLV- 1443
QY 226 YKNTI-IKTVAIPALSSGIF-----OEPNLICT-----KTIVETIRV 261
DB 1444 --NEHNIKSVAILPLSLTGIYVAGKDRLEVSINCLITLALDRTDADVTYICLDKKMERIDA 1501
QY 262 SLOGKPMMSNIKE 274
DB 1502 ALQKESVTELEKD 1514

```

# RESULT 11

US-09-102-248-9

Sequence 9, Application US/09102248

Patent No. 6008035

GENERAL INFORMATION:

APPLICANT: Johnston, Robert E.

APPLICANT: Davis, Nancy L.

APPLICANT: Simpson, Dennis A.

TITLE OF INVENTION: System for the In Vivo Delivery and

NUMBER OF SEQUENCES: 12 Expression of Heterologous Genes in the Bone Marrow

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bell Seltzer Park & Gibson, P.A.

STREET: 1211 East Morehead Street

CITY: Charlotte

STATE: No. 6008035th Carolina

COUNTRY: USA

ZIP: 28234

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/102,248

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/801,263

FILING DATE: 19-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: Sibley, Kenneth D.

REGISTRATION NUMBER: 31,665

REFERENCE/DOCKET NUMBER: 5470-147

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-420-2200

TELEFAX: 919-881-3175

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 2512 amino acids

TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-102-248-9

Query Match 3.0%; Score 132.5; DB 3; Length 2512;  
Best Local Similarity 23.0%; Pred. No. 0.021;  
Matches 72; Conservative 45; Mismatches 107; Indels 89; Gaps 13;

Query 26 FQKVAQIPQWRKGTCTCLPY-----KCEETGALGENTSWQIPINHNDFILKN----- 76  
DB 1227 YDLVFINICTKRNHHFQCCEDHAATLKLSSALN-----CLNPGGITLVKSKGYAD 1279  
QY 77 -NEROLCEVLQNKFGCISTLSPVQEGNSKSLQVERKM-----LTPR-----IELSVN- 123  
DB 1280 RNSIEDVTALARKFVNSAPDCVSSNTEMYLIFQLDINSRTQFTPHHLNCVISYVE 1339  
QY 124 -----KDDLTTHAVDAVYVNAANEDLLHGGGLALVAKAGFELQESKOF 168  
DB 1340 GTRDVGAAPSYRTKRENTADCOEEAVVNAANPLGRGEGVCRAIK----- 1386  
QY 169 VARYGVSAGEIAVTGAGRLP-C--KQIIHAVGPRMENDKOGCTGKLORAIVSILNYI 225  
DB 1387 -RWPSTFSDATETGTAMTYCLGKKVTHAVGDFRKHPEALKLLQNAVHAVADLV- 1443  
QY 226 YKNTH-IKVVAIPALSSGIF-----OPLNLCT-----KTIYETIV 261  
DB 1444 --NEHHIKVAIPLSTGIYAAGKRLLEVSLNCLFTALDRTADVYIYCLDKMKKERIDA 1501  
QY 262 SLOGKPMNSLKE 274  
DB 1502 ALQKESVTELD 1514

RESULT 12  
US-09-196-387-2  
Sequence 2, Application US/09196387  
Patent No. 6277613

GENERAL INFORMATION:  
APPLICANT: de Lange, Titia  
APPLICANT: Smith, Susan  
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TREP1 AND METHODS  
TITLE OF INVENTION: OF USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/196.387  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/095,225  
FILING DATE: June 10, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1327 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-196-387-2

Query Match 3.0%; Score 132; DB 4; Length 1327;  
Best Local Similarity 21.0%; Pred. No. 0.0081;  
Matches 65; Conservative 48; Mismatches 107; Indels 90; Gaps 15;

QY 558 EHDILSLOKTSSTVITELISPGTELEEGARA--DLIEVVMNIEDMLCKVOEEMARKK 615  
DB 1041 EH--LMDIETBQITLDVLADMGHEELKEIGINAVGHRKLIKGVRL-----CG 1089  
QY 616 ERG-----LMRLSGQMTI-----QQCKODEKKEVITFLKCPVPTQELDQKOF 661  
DB 1090 QOQTNEVLFHCVNOQTILLDLAPEDKEVYSEEE-----QSTIREHSDG 1135  
QY 662 ERGCG-----LOYLKVEKIDNEVLMMAFORKKKMEKELHROPVSHRLFQOVPEYFCNVYC 716  
DB 1136 GNAGGIFENYVNIIRIOKVYKKLRERFCHROKEVSENNHHNERNLFFGSP--FINALI 1193  
QY 717 RVGFORMISTPCDDPKYAGIYFTKLNLAERAKKISAA-----DKLIYFEAEVLTG 769  
DB 1194 HKGFEDRHAY-IGCMFGAGIYFAENSSKSNQVYVYGIGGTCGPTKDRSCYICHRQML-- 1250  
QY 770 FFCQ-----GHPLENIYPPPLSGAIDGHSVDVNS---SPETFYIFSG 810  
DB 1251 -FCRYTLGKSFLOFSYMKNAH---AP-----GHHSVIGRPSVNCGLAIAEVIYING 1297  
QY 811 MQAIPQYLTW 820  
DB 1298 EQAYPEYLT 1307

RESULT 13  
US-08-801-263A-5

Sequence 5, Application US/08801263A  
Patent No. 5811407

GENERAL INFORMATION:  
APPLICANT: Johnston, Robert E.  
APPLICANT: Davis, Nancy L.  
APPLICANT: Simpson, Dennis A.  
TITLE OF INVENTION: System for the In Vivo Delivery and  
TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bell Seltzer Park & Gibson, P.A.  
STREET: 1211 East Morehead Street  
CITY: Charlotte  
STATE: No. 5811407th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,263A  
FILING DATE: 19-FEB-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5470-147  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-420-2200  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ. ID NO. 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2517 amino acids

TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-801-263A-5

Query Match  
 Best Local Similarity 2.9%; Score 131; DB 2; Length 2517;  
 Matches 123; Conservative 85; Mismatches 216; Indels 208; Gaps 26;

QY 4 SMVAGAAVNEKSGR--ITSLSLFOKVAFOIFPOMKRGNTTECLPYKCSFTGALGENYS 61  
 DB 1269 TLVVKSYGADRNSEDVYALARKFVRVSA-ARPECVSNTTEMYLIFR-----QDINSRT 1322  
 QY 62 WQIPINHNDFILKNNEROLCEVLONKFCISTLVSPVOEGNSKSLQVFRKMLTPRIELS 121  
 DB 1323 RQFTPHH-----LNCVTSYVEGTRDVGGAAPSYRTKR----- 1355  
 QY 122 VKKDLTHAVDAVYVNAANEDLLHGGGLALVYKAGGFEIOESKOFVARGKVSAGETA 181  
 DB 1356 ---ENIADCOEAVVNAANPLGRPEGVCRAIYK-----RWFNSTDSAT 1397  
 QY 182 VTGAGRLP-C--KQIIHAVGPRMMEWDKOGCTGKLORAIVSLNVYIKNTH-IKTVAIP 237  
 DB 1398 ETGTAKLTVCGCKKVIHAAGPDRKHPREAEALKLLQNAHVAADLV---NEHNKISVAIP 1454  
 QY 238 ALSSGIF-----QFPLNCTKTIYETIRVSLQCKPMMSNLKEIHLVSNEDPTVAAP-K 289  
 DB 1455 LLSGTIYAAGKDRLEVSLLNCLTTALDRT-----DADVTTCYCLDK 1493  
 QY 290 AASEFILKSELGQETTPSFNAMYVNNLTLQIVOGHIEMQTADVIVNS---VNPHDITV 345  
 DB 1494 KWKERT-----DAVLQKESVTELDKEDMEIDELVWIHPDSCIK 1533  
 QY 346 GPVAKSILOQAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVHLMSE----- 399  
 DB 1534 G-----RKGSFTTKGKLY-----SYFEGTKRFQAAKDMAE 1563  
 QY 400 ---PPKPO-----ILKHAMKECLEKC-IEONITSI---SPPALGTGMELIKKEIA 442  
 DB 1564 IKVLEFPNDOESNEOLCAVILGETMAIREKCVVDHNPSSPPKTLPCLOMAYMTER--- 1620  
 QY 443 AEILFDEVLTFAKHVHQVLTKFYIFPTDLEIYKAFSSSEMAKRSKMLSLNNY----- 495  
 DB 1621 ---VHRLRSNNVK-EVT---VCSSTPLPKVIKIKNVQKQCVKVLVFNHPTAFVPA 1669  
 QY 496 -----SYPOSTREKRENGLEARSPAINLMGFVNEEYEAHAMIQRLISLQNHHT 545  
 DB 1670 RXYIAPQOPAPPAQAEAPVAAATPPPADNTSLDVTD-----ISLDMEDS 1718  
 QY 546 IENNHIILYGRKREHDILSLOKTSSVITEII 577  
 DB 1719 SEGSLFSSFGSDNSTITSMDSWSSGPPSSLEYI 1750

RESULT 14  
 US-09-102-248-5

Sequence 5, Application US/09102248  
 Patent No. 6008035

GENERAL INFORMATION:  
 APPLICANT: Johnston, Robert E.  
 APPLICANT: Davis, Nancy L.  
 APPLICANT: Simpson, Dennis A.  
 TITLE OF INVENTION: System for the In Vivo Delivery and  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Bell Seltzer Park & Gibson, P.A.  
 STREET: 1211 East Morehead Street  
 CITY: Charlotte  
 STATE: No. 6008035th Carolina  
 COUNTRY: USA  
 ZIP: 28234  
 COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/102,248  
 FILING DATE:  
 CLASSIFICATION:

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/801,263  
 FILING DATE: 19-FEB-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Sibley, Kenneth D.  
 REGISTRATION NUMBER: 31,665  
 REFERENCE/DOCKET NUMBER: 5470-147  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-420-2200  
 TELEFAX: 919-881-3175  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2517 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-102-248-5

Query Match  
 Best Local Similarity 2.9%; Score 131; DB 3; Length 2517;  
 Matches 123; Conservative 85; Mismatches 216; Indels 208; Gaps 26;

QY 4 SMVAGAAVNEKSGR--ITSLSLFOKVAFOIFPOMKRGNTTECLPYKCSFTGALGENYS 61  
 DB 1269 TLVVKSYGADRNSEDVYALARKFVRVSA-ARPECVSNTTEMYLIFR-----QDINSRT 1322  
 QY 62 WQIPINHNDFILKNNEROLCEVLONKFCISTLVSPVOEGNSKSLQVFRKMLTPRIELS 121  
 DB 1323 RQFTPHH-----LNCVTSYVEGTRDVGGAAPSYRTKR----- 1355  
 QY 122 VKKDLTHAVDAVYVNAANEDLLHGGGLALVYKAGGFEIOESKOFVARGKVSAGETA 181  
 DB 1356 ---ENIADCOEAVVNAANPLGRPEGVCRAIYK-----RWFNSTDSAT 1397  
 QY 182 VTGAGRLP-C--KQIIHAVGPRMMEWDKOGCTGKLORAIVSLNVYIKNTH-IKTVAIP 237  
 DB 1398 ETGTAKLTVCGCKKVIHAAGPDRKHPREAEALKLLQNAHVAADLV---NEHNKISVAIP 1454  
 QY 238 ALSSGIF-----QFPLNCTKTIYETIRVSLQCKPMMSNLKEIHLVSNEDPTVAAP-K 289  
 DB 1455 LLSGTIYAAGKDRLEVSLLNCLTTALDRT-----DADVTTCYCLDK 1493  
 QY 290 AASEFILKSELGQETTPSFNAMYVNNLTLQIVOGHIEMQTADVIVNS---VNPHDITV 345  
 DB 1494 KWKERT-----DAVLQKESVTELDKEDMEIDELVWIHPDSCIK 1533  
 QY 346 GPVAKSILOQAGVEMKSEFLATKAKOFORSOLVLTGKFNLFCKYIYVHLMSE----- 399  
 DB 1534 G-----RKGSFTTKGKLY-----SYFEGTKRFQAAKDMAE 1563  
 QY 400 ---PPKPO-----ILKHAMKECLEKC-IEONITSI---SPPALGTGMELIKKEIA 442  
 DB 1564 IKVLEFPNDOESNEOLCAVILGETMAIREKCVVDHNPSSPPKTLPCLOMAYMTER--- 1620  
 QY 443 AEILFDEVLTFAKHVHQVLTKFYIFPTDLEIYKAFSSSEMAKRSKMLSLNNY----- 495  
 DB 1621 ---VHRLRSNNVK-EVT---VCSSTPLPKVIKIKNVQKQCVKVLVFNHPTAFVPA 1669  
 QY 496 -----SYPOSTREKRENGLEARSPAINLMGFVNEEYEAHAMIQRLISLQNHHT 545  
 DB 1670 RXYIAPQOPAPPAQAEAPVAAATPPPADNTSLDVTD-----ISLDMEDS 1718  
 QY 546 IENNHIILYGRKREHDILSLOKTSSVITEII 577  
 DB 1719 SEGSLFSSFGSDNSTITSMDSWSSGPPSSLEYI 1750







GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:44:37 ; Search time 56 Seconds  
(without alignments)  
1403.392 Million cell updates/sec

Title: US-09-830-762-2  
Perfect score: 4447  
Sequence: 1 MDFSVAAGAAAYNEKSGRIT.....PMRPFQHPMRGFAAGSPVD 854

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*

2: /cgn2\_6/ptodata/2/pubppaa/PCIT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*

6: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/2/pubppaa/PCITUS\_PUBCOMB.pep:\*

8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*

10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*

12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*

13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*

14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description        |
|------------|--------|-------------|--------|-------|--------------------|
| 1          | 4393   | 98.8        | 856    | 10    | US-09-882-529-4    |
| 2          | 4208.5 | 94.6        | 819    | 10    | US-09-882-529-2    |
| 3          | 4175.5 | 93.9        | 821    | 10    | US-09-882-529-3    |
| 4          | 912    | 20.5        | 179    | 10    | US-09-882-529-7    |
| 5          | 841.5  | 18.9        | 169    | 10    | US-09-882-529-5    |
| 6          | 751    | 16.9        | 145    | 10    | US-09-882-529-9    |
| 7          | 664    | 14.9        | 129    | 10    | US-09-882-529-15   |
| 8          | 626    | 14.1        | 121    | 10    | US-09-882-529-11   |
| 9          | 626    | 14.1        | 121    | 10    | US-09-882-529-13   |
| 10         | 244.5  | 5.5         | 170    | 10    | US-09-882-529-6    |
| 11         | 244.5  | 5.5         | 170    | 10    | US-09-882-529-8    |
| 12         | 215.5  | 4.8         | 132    | 10    | US-09-882-529-10   |
| 13         | 215.5  | 4.8         | 132    | 10    | US-09-882-529-16   |
| 14         | 179.5  | 4.0         | 116    | 10    | US-09-882-529-12   |
| 15         | 179.5  | 4.0         | 116    | 10    | US-09-882-529-14   |
| 16         | 157    | 3.5         | 378    | 10    | US-09-925-301-1208 |
| 17         | 156    | 3.5         | 204    | 9     | US-09-731-001-5    |
| 18         | 156    | 3.5         | 220    | 9     | US-09-731-001-2    |
| 19         | 156    | 3.5         | 258    | 9     | US-09-731-001-4    |

|    |       |     |      |    |                     |                    |
|----|-------|-----|------|----|---------------------|--------------------|
| 20 | 156   | 3.5 | 598  | 9  | US-09-731-001-3     | Sequence 3, Appl1  |
| 21 | 148   | 3.3 | 716  | 10 | US-09-815-242-12403 | Sequence 12403, A  |
| 22 | 143   | 3.2 | 2492 | 10 | US-09-991-258-3     | Sequence 3, Appl1  |
| 23 | 140.5 | 3.2 | 250  | 9  | US-10-102-806-517   | Sequence 517, App  |
| 24 | 137   | 3.1 | 1166 | 9  | US-09-972-115A-6    | Sequence 6, Appl1  |
| 25 | 134   | 3.0 | 802  | 9  | US-09-964-899-41    | Sequence 41, Appl1 |
| 26 | 134   | 3.0 | 1074 | 10 | US-09-509-196A-2    | Sequence 2, Appl1  |
| 27 | 133   | 3.0 | 1786 | 9  | US-09-742-096-3     | Sequence 8, Appl1  |
| 28 | 132   | 3.0 | 1327 | 10 | US-09-972-115A-8    | Sequence 2, Appl1  |
| 29 | 132   | 3.0 | 1327 | 10 | US-09-841-835-2     | Sequence 2, Appl1  |
| 30 | 132   | 3.0 | 1945 | 9  | US-09-927-587-2     | Sequence 4, Appl1  |
| 31 | 132   | 3.0 | 1979 | 9  | US-09-927-587-4     | Sequence 4, Appl1  |
| 32 | 130.5 | 2.9 | 465  | 10 | US-09-815-242-5335  | Sequence 5335, Ap  |
| 33 | 130   | 2.9 | 374  | 10 | US-10-091-483-236   | Sequence 236, App  |
| 34 | 130   | 2.9 | 374  | 10 | US-09-764-046-236   | Sequence 236, App  |
| 35 | 124.5 | 2.8 | 1019 | 10 | US-09-801-574-76    | Sequence 76, Appl  |
| 36 | 124   | 2.8 | 2139 | 9  | US-10-023-219-4     | Sequence 4, Appl1  |
| 37 | 124   | 2.8 | 2139 | 10 | US-09-727-384-6     | Sequence 6, Appl1  |
| 38 | 123.5 | 2.8 | 822  | 10 | US-09-981-900B-9    | Sequence 9, Appl1  |
| 39 | 123.5 | 2.8 | 1333 | 9  | US-09-972-115A-2    | Sequence 2, Appl1  |
| 40 | 122.5 | 2.8 | 1267 | 9  | US-09-972-115A-4    | Sequence 4, Appl1  |
| 41 | 122   | 2.7 | 846  | 9  | US-09-893-519A-34   | Sequence 34, Appl  |
| 42 | 117.5 | 2.6 | 617  | 10 | US-09-815-242-13228 | Sequence 13228, A  |
| 43 | 117   | 2.6 | 2012 | 9  | US-09-808-602-68    | Sequence 68, Appl  |
| 44 | 116   | 2.6 | 2037 | 9  | US-09-951-402-3     | Sequence 3, Appl1  |
| 45 | 116   | 2.6 | 2037 | 10 | US-09-951-401-3     | Sequence 3, Appl1  |

ALIGNMENTS

RESULT 1  
US-09-882-529-4  
Sequence 4, Application US/09882529  
Patent No. US20020132317A1  
GENERAL INFORMATION:  
APPLICANT: Peyman, John A  
APPLICANT: da Silva, Antonio  
TITLE OF INVENTION: Hockman, Paula  
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
FILE REFERENCE: 15966-771  
CURRENT APPLICATION NUMBER: US/09/882, 529  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: 60/211, 565  
PRIOR FILING DATE: 2000-06-15  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 856  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-882-529-4

|                       |              |  |               |             |
|-----------------------|--------------|--|---------------|-------------|
| Query Match           | 98.8%        | Score 4393;  | DB 10;        | Length 856; |
| Best Local Similarity | 99.2%        | Pred. No. 0;   |               |             |
| Matches 849;          | Conservative | 0;   | Mismatches 5; | Indels 2;   |
| Gaps 2;               |              |  |               |             |
| QY                    | 1            | MDFSVAAGAAAYNEKSGRITSLILFOKVFNAIFPQWKGNTFEDLPKCSGTGALGNY 60  |               |             |
| DB                    | 1            | MDFSVAAGAAAYNEKSGRITSLILFOKVFNAIFPQWKGNTFEDLPKCSGTGALGNY 60  |               |             |
| QY                    | 61           | SMQIINNDFFILKNNEHOLCEVONKFGCSTLVSPOEBSKSLQVFRKMLPRIEL 120    |               |             |
| DB                    | 61           | SMQIINNDFFILKNNEHOLCEVONKFGCSTLVSPOEBSKSLQVFRKMLPRIEL 120    |               |             |
| QY                    | 121          | SVWKDILTTTHAVDAVYVNAEDLHGGGLALALVKAAGFEIOESKQFVARYGSAGEI 180 |               |             |
| DB                    | 121          | SVWKDILTTTHAVDAVYVNAEDLHGGGLALALVKAAGFEIOESKQFVARYGSAGEI 180 |               |             |
| QY                    | 181          | AVTGAGRLPCKOIITHAVGRMWMKQCTGKLOAIVSILNYVYKTHIKTVAIPALS 240   |               |             |
| DB                    | 181          | AVTGAGRLPCKOIITHAVGRMWMKQCTGKLOAIVSILNYVYKTHIKTVAIPALS 240   |               |             |

```

QY 241 SGIFPPLNLCTKTIVETIRVSLQCKPMMNSLKEIHLVSNEDPTVAARKASEFTLGKSE 300
DB 241 SGIFPPLNLCTKTIVETIRVSLQCKPMMNSLKEIHLVSNEDPTVAARKASEFTLGKSE 300
QY 301 LGOETTPSFNMAVNNLTLOIVQGHIEWOTADVIYNSVNPBDITVGPVAKSILQOAGVEM 360
DB 301 LGOETTPSFNMAVNNLTLOIVQGHIEWOTADVIYNSVNPBDITVGPVAKSILQOAGVEM 360
QY 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
DB 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
QY 421 NITSISFPALGNGNEIKETAEITLFDVLTFF-AKDHVKHQLTYKVFIFPDLEIYKAF 479
DB 421 NITSISFPALGNGNEIKETAEITLFDVLTFF-AKDHVKHQLTYKVFIFPDLEIYKAF 480
QY 480 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSPTAINTMGFNVEEMVFAHAMTORILIS 539
DB 481 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSPTAINTMGFNVEEMVFAHAMTORILIS 540
QY 540 LONHHIITNNHIIYLGKREHDLISOLQTSVSTIEIISPGTELEIEGARADLIEVYNN 599
DB 541 LONHHIITNNHIIYLGKREHDLISOLQTSVSTIEIISPGTELEIEGARADLIEVYNN 600
QY 600 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 659
DB 601 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 660
QY 660 QFEKCGLOVLKVEKIDNEVYLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 719
DB 661 QFEKCGLOVLKVEKIDNEVYLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 720
QY 720 FORMSTPCDDPKYAGIYFTTNKLNLAERAKKISADKLIVFEAEVLTGFFCOGHPPLNT 779
DB 721 FORMSTPCDDPKYAGIYFTTNKLNLAERAKKISADKLIVFEAEVLTGFFCOGHPPLNT 780
QY 780 VPPPLSPGALDGHDSVDVNVSSPETF-VIFSGMOAIPQYLTCTQOEVYQSDYSSGPMRPA 838
DB 781 VPPPLSPGALDGHDSVDVNVSSPETF-VIFSGMOAIPQYLTCTQOEVYQSDYSSGPMRPA 840
QY 839 FAQHPWRGFASSGSPVD 854
DB 841 FAQHPWRGFASSGSPVD 856

```

RESULT 2

US-09-882-529-2

Sequence 2, Application US/09882529

Patent No. US20020132317A1

GENERAL INFORMATION:

APPLICANT: Peyman, John A

APPLICANT: da Silva, Antonio

APPLICANT: Hockman, Paula

TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC

FILE REFERENCE: 15966-771

CURRENT APPLICATION NUMBER: US/09/882, 529

PRIOR FILING DATE: 2001-09-12

PRIOR APPLICATION NUMBER: 60/211, 565

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 819

TYPE: PRT

ORGANISM: Homo sapiens

US-09-882-529-2

Query Match 94.6%; Score 4208.5; DB 10; Length 819;

Best Local Similarity 95.4%; Pred. No. 0;

Matches 815; Conservative 3; Mismatches 1; Indels 35; Gaps 1;

```

QY 1 MDFSVAAGAAVNEKSGRTISLSLLFQKVFQAIFFQMKRGKNTTECLPYKCSSETGALGENY 60
DB 1 MDFSVAAGAAVNEKSGRTISLSLLFQKVFQAIFFQMKRGKNTTECLPYKCSSETGALGENY 60
QY 61 SMQIDPNNHNDFFILKNNEKROCEVLONKFGCISTLYSPVOEGNSKSLQVFRMLLPRIEL 120
DB 61 SMQIDPNNHNDFFILKNNEKROCEVLONKFGCISTLYSPVOEGNSKSLQVFRMLLPRIEL 120
QY 121 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 180
DB 121 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 180
QY 180 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 145
DB 181 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 145
QY 145 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 145
DB 146 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 145
QY 146 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 145
DB 146 SVMKDGLTTHAVADVANNANEDLHGGGLALALVAGGFEIOEESKQFVARGYKASAEI 145
QY 241 SGIFPPLNLCTKTIVETIRVSLQCKPMMNSLKEIHLVSNEDPTVAARKASEFTLGKSE 300
DB 241 SGIFPPLNLCTKTIVETIRVSLQCKPMMNSLKEIHLVSNEDPTVAARKASEFTLGKSE 300
QY 301 LGOETTPSFNMAVNNLTLOIVQGHIEWOTADVIYNSVNPBDITVGPVAKSILQOAGVEM 360
DB 301 LGOETTPSFNMAVNNLTLOIVQGHIEWOTADVIYNSVNPBDITVGPVAKSILQOAGVEM 360
QY 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
DB 361 KSEFLATKAKOFORSOLVLTGKFNLFCKYIYHVLHSEFPKPOLIKHAMKECLEKCIQ 420
QY 421 NITSISFPALGNGNEIKETAEITLFDVLTFF-AKDHVKHQLTYKVFIFPDLEIYKAF 479
DB 421 NITSISFPALGNGNEIKETAEITLFDVLTFF-AKDHVKHQLTYKVFIFPDLEIYKAF 480
QY 480 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSPTAINTMGFNVEEMVFAHAMTORILIS 539
DB 481 SSEMAKRSKMLSLNNYSVQOSTREKRENGLEARSPTAINTMGFNVEEMVFAHAMTORILIS 540
QY 540 LONHHIITNNHIIYLGKREHDLISOLQTSVSTIEIISPGTELEIEGARADLIEVYNN 599
DB 541 LONHHIITNNHIIYLGKREHDLISOLQTSVSTIEIISPGTELEIEGARADLIEVYNN 600
QY 600 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 659
DB 601 IEDMLCKVOEEMARKKERGLMRSLGOWTIOOQKTODEKKNITFLKCPVPPTOELLDOKK 660
QY 660 QFEKCGLOVLKVEKIDNEVYLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 719
DB 661 QFEKCGLOVLKVEKIDNEVYLAFAFORKKKMEKLRQPVSHRLEQOVPYQFCNVVCRVGF 720
QY 720 FORMSTPCDDPKYAGIYFTTNKLNLAERAKKISADKLIVFEAEVLTGFFCOGHPPLNT 779
DB 721 FORMSTPCDDPKYAGIYFTTNKLNLAERAKKISADKLIVFEAEVLTGFFCOGHPPLNT 780
QY 780 VPPPLSPGALDGHDSVDVNVSSPETF-VIFSGMOAIPQYLTCTQOEVYQSDYSSGPMRPA 838
DB 781 VPPPLSPGALDGHDSVDVNVSSPETF-VIFSGMOAIPQYLTCTQOEVYQSDYSSGPMRPA 840
QY 841 QHPWRGFASSGSPVD 854
DB 841 QHPWRGFASSGSPVD 856

```

RESULT 3

US-09-882-529-3

Sequence 3, Application US/09882529

Patent No. US20020132317A1

GENERAL INFORMATION:

APPLICANT: Peyman, John A

APPLICANT: da Silva, Antonio

APPLICANT: Hockman, Paula

TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC

FILE REFERENCE: 15966-771

CURRENT APPLICATION NUMBER: US/09/882, 529

; CURRENT FILING DATE: 2001-09-12  
 ; PRIOR APPLICATION NUMBER: 60/211,565  
 ; PRIOR FILING DATE: 2000-06-15  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 821  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-882-529-3

Query Match 93.9%; Score 4175.5; DB 10; Length 821;  
 Best Local Similarity 95.1%; Pred. No. 0;  
 Matches 814; Conservative 0; Mismatches 5; Indels 37; Gaps 3;

QY 1 MPFSWAGAAAYNEKSGRTISLSLFQKFAQIFPQWRKGNTECLPYCSGALGENY 60  
 DB 1 MPFSWAGAAAYNEK-----SETGALGENY 25  
 QY 61 SMOIPINNDPFLIKKNERQCEVLONKPGCISTLVSPOEGNSKSLQVFRKKLTFRTEL 120  
 DB 26 SMOIPINNDPFLIKKNERQCEVLONKPGCISTLVSPOEGNSKSLQVFRKKLTFRTEL 85  
 QY 121 SYWKDDLTTHAVDAVVAANEDLLHGGGLALALVAGGFELQESKQFVARYGKSAGEI 180  
 DB 86 SYWKDDLTTHAVDAVVAANEDLLHGGGLALALVAGGFELQESKQFVARYGKSAGEI 145  
 QY 181 ATYGAGRLCKOIHAHVGRMWMKOGCTGKLOAIVSLNVIYKNTHTVAIPALPS 240  
 DB 146 ATYGAGRLCKOIHAHVGRMWMKOGCTGKLOAIVSLNVIYKNTHTVAIPALPS 205  
 QY 241 SGIFQPLNLCRTIVETIRVSLQGRPMNSLKEIHLVSNEDPTVAAFKASEFLLGKSE 300  
 DB 206 SGIFQPLNLCRTIVETIRVSLQGRPMNSLKEIHLVSNEDPTVAAFKASEFLLGKSE 265  
 QY 301 LQOETPSTNANVNNLTQIYQGHLENOTADYVNSVPHDITVGPVAKSIIQOAGVEM 360  
 DB 266 LQOETPSTNANVNNLTQIYQGHLENOTADYVNSVPHDITVGPVAKSIIQOAGVEM 325  
 QY 361 KSEFLATKAKOFORSOLVLTGKGFNLCKYIYHVLHSEFPKPOLKHKMKECELCIO 420  
 DB 326 KSEFLATKAKOFORSOLVLTGKGFNLCKYIYHVLHSEFPKPOLKHKMKECELCIO 385  
 QY 421 NITSISFPALGTGNMEIKETAELFDEVLTFF-ANDVYKHOLTVKVFJPTDLEIYKAF 479  
 DB 386 NITSISFPALGTGNMEIKETAELFDEVLTFF-ANDVYKHOLTVKVFJPTDLEIYKAF 445  
 QY 480 SSEMARSKMLSLNNSVQSTREKRENGLEKSPAINLMGFNVSENEEAHAMIORILS 539  
 DB 446 SSEMARSKMLSLNNSVQSTREKRENGLEKSPAINLMGFNVSENEEAHAMIORILS 505  
 QY 540 LONHHTIENNHLIYLGRRKHDILSOLOKTSSVITELISPGFTELEIEGARADILEVNA 599  
 DB 506 LONHHTIENNHLIYLGRRKHDILSOLOKTSSVITELISPGFTELEIEGARADILEVNA 565  
 QY 600 IEDMLCKVOEBARKKERGLMSLQGTIOOQKTODEMKNILFLKCPVPPTOLLDOKK 659  
 DB 566 IEDMLCKVOEBARKKERGLMSLQGTIOOQKTODEMKNILFLKCPVPPTOLLDOKK 625  
 QY 660 OFEKKGLQVLYKVEKIDNEVLMAAFQRRKKMEKELHROVSHRLFOOVYQFCNVYCRG 719  
 DB 626 OFEKKGLQVLYKVEKIDNEVLMAAFQRRKKMEKELHROVSHRLFOOVYQFCNVYCRG 685  
 QY 720 FORMYSTPCDPKYGAGIYFTKLNKLAERAKKISADKLIYVEAEVLITGFCOGHPLNT 779  
 DB 686 FORMYSTPCDPKYGAGIYFTKLNKLAERAKKISADKLIYVEAEVLITGFCOGHPLNT 745  
 QY 780 VPPPLSPGALIDGHDVAVNVSSPEFV-VIFSGMOAIPQYLTMTCTOEVVOSQDYSSGPMRP 838  
 DB 746 VPPPLSPGALIDGHDVAVNVSSPEFV-VIFSGMOAIPQYLTMTCTOEVVOSQDYSSGPMRP 805  
 QY 839 FAQHPWRGFASSGSPVD 854  
 ||||||||||||||||

DB 806 FAQHPWRGFASSGSPVD 821

RESULT 4  
 US-09-882-529-7  
 ; Sequence 7, Application US/09882529  
 ; Patent No. US20020132317A1  
 ; GENERAL INFORMATION:

; APPLICANT: Peyman, John A  
 ; APPLICANT: da Silva, Antonio  
 ; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
 ; FILE REFERENCE: 15966-771  
 ; CURRENT APPLICATION NUMBER: US/09/882,529  
 ; CURRENT FILING DATE: 2001-09-12  
 ; PRIOR APPLICATION NUMBER: 60/211,565  
 ; PRIOR FILING DATE: 2000-06-15  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 7  
 ; LENGTH: 179  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-882-529-7

Query Match 20.5%; Score 912; DB 10; Length 179;  
 Best Local Similarity 99.4%; Pred. No. 2,7e-64;  
 Matches 178; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 114 LTPRIELSYWKDDLTTHAVDAVVAANEDLLHGGGLALALVAGGFELQESKQFVARYG 173  
 DB 1 LTPRIELSYWKDDLTTHAVDAVVAANEDLLHGGGLALALVAGGFELQESKQFVARYG 60  
 QY 174 KVSAGEIATYGAGRLCKOIHAHVGRMWMKOGCTGKLOAIVSLNVIYKNTHT 233  
 DB 61 KVSAGEIATYGAGRLCKOIHAHVGRMWMKOGCTGKLOAIVSLNVIYKNTHT 120  
 QY 234 VAIPALSSGIFQPLNLCRTIVETIRVSLQGRPMNSLKEIHLVSNEDPTVAAFKAS 292  
 DB 121 VAIPALSSGIFQPLNLCRTIVETIRVSLQGRPMNSLKEIHLVSNEDPTVAAFKAS 179

RESULT 5  
 US-09-882-529-5  
 ; Sequence 5, Application US/09882529  
 ; Patent No. US20020132317A1  
 ; GENERAL INFORMATION:

; APPLICANT: Peyman, John A  
 ; APPLICANT: da Silva, Antonio  
 ; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
 ; FILE REFERENCE: 15966-771  
 ; CURRENT APPLICATION NUMBER: US/09/882,529  
 ; CURRENT FILING DATE: 2001-09-12  
 ; PRIOR APPLICATION NUMBER: 60/211,565  
 ; PRIOR FILING DATE: 2000-06-15  
 ; NUMBER OF SEQ ID NOS: 16  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 169  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-882-529-5

Query Match 18.9%; Score 841.5; DB 10; Length 169;  
 Best Local Similarity 97.6%; Pred. No. 9.1e-59;  
 Matches 165; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  
 QY 313 VVNNLTLOIVQGHIEQMTADYVNSVPHDITVGPVAKSIIQOAGVEMKSEFLATKAKOF 372  
 DB 1 VVNNLTLOIVQGHIEQMTADYVNSVPHDITVGPVAKSIIQOAGVEMKSEFLATKAKOF 60

Query Match 14.9%; Score 751; DB 10; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1e-51;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 111 RKMLPRLELSWKDDLTTHAVDAVYNAANEDLHGGLALALVYAGGFEIOESKOFYA 170  
Db 1 RKMLPRLELSWKDDLTTHAVDAVYNAANEDLHGGLALALVYAGGFEIOESKOFYA 60  
Qy 171 RRGKSAGSIAYTGAGRLPCKOIIHAVGPRMWMKOGCTGLQRAIVSILNVYTKNTH 230  
Db 61 RRGKSAGSIAYTGAGRLPCKOIIHAVGPRMWMKOGCTGLQRAIVSILNVYTKNTH 120  
Qy 231 IKTVAI PALSSGIFQFPPLNCTKTI 255  
Db 121 IKTVAI PALSSGIFQFPPLNCTKTI 145

RESULT 6  
US-09-882-529-9  
Sequence 9, Application US/09882529  
Patent No. US20020132317A1  
GENERAL INFORMATION:  
APPLICANT: Peyman, John A  
APPLICANT: da Silva, Antonio  
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
FILE REFERENCE: 15966-771  
CURRENT APPLICATION NUMBER: US/09/882,529  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: 60/211,565  
PRIOR FILING DATE: 2000-06-15  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 145  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-882-529-9

Query Match 16.9%; Score 751; DB 10; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1e-51;  
Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 111 RKMLPRLELSWKDDLTTHAVDAVYNAANEDLHGGLALALVYAGGFEIOESKOFYA 170  
Db 1 RKMLPRLELSWKDDLTTHAVDAVYNAANEDLHGGLALALVYAGGFEIOESKOFYA 60  
Qy 171 RRGKSAGSIAYTGAGRLPCKOIIHAVGPRMWMKOGCTGLQRAIVSILNVYTKNTH 230  
Db 61 RRGKSAGSIAYTGAGRLPCKOIIHAVGPRMWMKOGCTGLQRAIVSILNVYTKNTH 120  
Qy 231 IKTVAI PALSSGIFQFPPLNCTKTI 255  
Db 121 IKTVAI PALSSGIFQFPPLNCTKTI 145

RESULT 7  
US-09-882-529-15  
Sequence 15, Application US/09882529  
Patent No. US20020132317A1  
GENERAL INFORMATION:  
APPLICANT: Peyman, John A  
APPLICANT: da Silva, Antonio  
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
FILE REFERENCE: 15966-771  
CURRENT APPLICATION NUMBER: US/09/882,529  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: 60/211,565  
PRIOR FILING DATE: 2000-06-15  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 15  
LENGTH: 129  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-882-529-15

Query Match 14.9%; Score 664; DB 10; Length 129;  
Best Local Similarity 99.2%; Pred. No. 6.1e-45;  
Matches 128; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Query 318 TLQIVQGHLEWMTADVYNSVNPHTITGVPVAKSILLOQAGVYMKSEFLATKAKOFORSOL 377  
Db 1 TLQIVQGHLEWMTADVYNSVNPHTITGVPVAKSILLOQAGVYMKSEFLATKAKOFORSOL 60  
Qy 378 VLVATGPNLFCKYIYHVLHMSSEFPKPOLKHAMKECLEKCIQNTSISFPALGTGNNEI 437  
Db 61 VLVATGPNLFCKYIYHVLHMSSEFPKPOLKHAMKECLEKCIQNTSISFPALGTGNNEI 120  
Qy 438 KKEIAAEIL 446  
Db 121 KKEIAAEIL 129

RESULT 8  
US-09-882-529-11  
Sequence 11, Application US/09882529  
Patent No. US20020132317A1  
GENERAL INFORMATION:  
APPLICANT: Peyman, John A  
APPLICANT: da Silva, Antonio  
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
FILE REFERENCE: 15966-771  
CURRENT APPLICATION NUMBER: US/09/882,529  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: 60/211,565  
PRIOR FILING DATE: 2000-06-15  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 11  
LENGTH: 121  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-882-529-11

Query Match 14.1%; Score 626; DB 10; Length 121;  
Best Local Similarity 100.0%; Pred. No. 5.5e-42;  
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 136 VNAANEDLHGGLALALVYKAGGFEIOESKOFVARYKVSAGETAVTGAGRLPCKOIIH 195  
Db 1 VNAANEDLHGGLALALVYKAGGFEIOESKOFVARYKVSAGETAVTGAGRLPCKOIIH 60  
Qy 196 AVGPRMWMKOGCTGLQRAIVSILNVYTKNTHIKTVAI PALSSGIFQFPPLNCTKTI 255  
Db 61 AVGPRMWMKOGCTGLQRAIVSILNVYTKNTHIKTVAI PALSSGIFQFPPLNCTKTI 120  
Qy 256 V 256  
Db 121 V 121

RESULT 9  
US-09-882-529-13  
Sequence 13, Application US/09882529  
Patent No. US20020132317A1  
GENERAL INFORMATION:  
APPLICANT: Peyman, John A  
APPLICANT: da Silva, Antonio  
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC  
FILE REFERENCE: 15966-771  
CURRENT APPLICATION NUMBER: US/09/882,529  
CURRENT FILING DATE: 2001-09-12  
PRIOR APPLICATION NUMBER: 60/211,565  
PRIOR FILING DATE: 2000-06-15  
NUMBER OF SEQ ID NOS: 16

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-13
```

```
Query Match
Best Local Similarity 14.1%; Score 626; DB 10; Length 121;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 136 VNAANEDLHGGLALALVAKAGFEIOESKQFVARYGVSAEIVATGAGRLPCKQIIT 195
DB 1 VNAANEDLHGGLALALVAKAGFEIOESKQFVARYGVSAEIVATGAGRLPCKQIIT 60
OY 196 AVGPFRMMDKQCGCTGKQRAIVSILNVIYKNTHTKTAIPALSSGIFQPLNCTKT 255
DB 61 AVGPFRMMDKQCGCTGKQRAIVSILNVIYKNTHTKTAIPALSSGIFQPLNCTKT 120
OY 256 V 256
DB 121 V 121
```

```
RESULT 10
US-09-882-529-6
; Sequence 6, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-6
```

```
Query Match
Best Local Similarity 5.5%; Score 244.5; DB 10; Length 170;
Matches 64; Conservative 31; Mismatches 70; Indels 11; Gaps 4;
```

```
OY 115 TPRIELSVKDDLTTHAVDAVNAANEDLHGGLALALVAKAGF-EIOESKQFVARYG 173
DB 2 TGSMTKIKVKGDTITKLPAADAVNAANSDLTMGGVAGAIARAAGEPELEEE---LKG 57
OY 174 KVSAGEIATVAGARLPCKQIITHAVGPRMMDKQCGCTGKQRAIVSILNVIYKNTHTKT 233
DB 58 GVPTEGAVVTPGGNLPKAKYVIAHAVGPRMMDKQCGCTGKQRAIVSILNVIYKNTHTKT 115
OY 234 VAIPALSSGIFQPLNCTKTIVETIRVSLQCKPMSNLKEIHLVSNEDPTVAAPK 289
DB 116 VAIPALSSGIFQPLNCTKTIVETIRVSLQCKPMSNLKEIHLVSNEDPTVAAPK 167
```

```
RESULT 11
US-09-882-529-8
; Sequence 8, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; APPLICANT: Hockman, Paula
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
```

```
; TITLE OF INVENTION: ACIDS ENCODING SAME
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-8
```

```
Query Match
Best Local Similarity 5.5%; Score 244.5; DB 10; Length 170;
Matches 64; Conservative 31; Mismatches 70; Indels 11; Gaps 4;
```

```
OY 115 TPRIELSVKDDLTTHAVDAVNAANEDLHGGLALALVAKAGF-EIOESKQFVARYG 173
DB 2 TGSMTKIKVKGDTITKLPAADAVNAANSDLTMGGVAGAIARAAGEPELEEE---LKG 57
OY 174 KVSAGEIATVAGARLPCKQIITHAVGPRMMDKQCGCTGKQRAIVSILNVIYKNTHTKT 233
DB 58 GVPTEGAVVTPGGNLPKAKYVIAHAVGPRMMDKQCGCTGKQRAIVSILNVIYKNTHTKT 115
OY 234 VAIPALSSGIFQPLNCTKTIVETIRVSLQCKPMSNLKEIHLVSNEDPTVAAPK 289
DB 116 VAIPALSSGIFQPLNCTKTIVETIRVSLQCKPMSNLKEIHLVSNEDPTVAAPK 167
```

```
RESULT 12
US-09-882-529-10
; Sequence 10, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 60/211,565
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-10
```

```
Query Match
Best Local Similarity 4.8%; Score 215.5; DB 10; Length 132;
Matches 53; Conservative 18; Mismatches 52; Indels 5; Gaps 3;
```

```
OY 120 LSWMKDDLTTHAVDAVNAANEDLHGGLALALVAKAGFEIOESKQFVARYGKVSAGE 179
DB 2 LKVKAGDITKPRADAVNAANSDDAHGGVAGAIARAAGE--ESKEEFKRLAGECPVGT 59
OY 180 IATGAGRLPCKQIITHAVGPRMMDKQCGCTGKQRAIVSILNVIYKNTHTKTAIPAL 239
DB 60 AVTEGAGNLPKAKYVIAHAVGPRMMDKQCGCTGKQRAIVSILNVIYKNTHTKTAIPAL 116
OY 240 SSGIFQFP 247
DB 117 GTGTYGVP 124
```

```
RESULT 13
US-09-882-529-16
```

```

Sequence 16: Application US/09862529
Patent No. US20020132317A1
GENERAL INFORMATION:
APPLICANT: Peyman, John A
APPLICANT: da Silva, Antonio
APPLICANT: Hockman, Paula
TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING SAME
FILE REFERENCE: 15966-771
CURRENT APPLICATION NUMBER: US/09/882,529
CURRENT FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: 60/211,565
PRIOR FILING DATE: 2000-06-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 132
TYPE: PRT
ORGANISM: Homo sapiens
US-09-862-529-16

```

|                       |       |   |       |                                 |
|-----------------------|-------|---|-------|---------------------------------|
| Query Match           | 4.8%  | Score 215.5   | DB 10 | Length 132                      |
| Best Local Similarity | 41.4% | Pred. No. 1.6e-09   |       |                                 |
| Matches               | 53    | Conservative  | 18    | Mismatches 52; Indels 5; Gaps 3 |
| QY                    | 120   | LSWKDDLTTHAVDAVVAANAEDLLHGGGLALALVKAGGFELQESKQFVARYGRVSAGE      | 179   |                                 |
|                       |       |   |       |                                 |
| Db                    | 2     | LKYVGGDITTKPRADIAIVNANSDGAHGSHVACALARAAGWE--EKEKEERKLAGECPVOT   | 59    |                                 |
|                       |       |   |       |                                 |
| QY                    | 180   | IATVAGAGFLPCKQIITHAVGPRMMEKDQCGCTGKLORAIYSIIINVIYKNTHTIKYVAIPAL | 239   |                                 |
|                       |       |   |       |                                 |
| Db                    | 60    | AVVEGGMLPKPKYVIAHVGPEASGYSKEGYE-LLENAYRACRIAI--ELGIKSAIPII      | 116   |                                 |
|                       |       |   |       |                                 |
| QY                    | 240   | SSGIQFP   | 247   |                                 |
|                       |       |   |       |                                 |
| Db                    | 117   | GTGITVGP  | 124   |                                 |

```

RESULT 14
US-09-882-529-12
; Sequence 12, Application US/09882529
; Patent No. US20020132317A1
; GENERAL INFORMATION:
; APPLICANT: Peyman, John A
; APPLICANT: da Silva, Antonio
; APPLICANT: Hockman, Paula
; TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
; FILE REFERENCE: 15966-771
; CURRENT APPLICATION NUMBER: US/09/882,529
; CURRENT FILING DATE: 2001-09-12
; PRIORITY APPLICATION NUMBER: 60/211,565
; PRIORITY FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 116
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-882-529-12

```

[illegible]

|    |       |     |
|----|-------|-----|
| QY | 256 V | 256 |
| Db | 116 L | 116 |

```

RESULT 15
US-09-882-529-14
: Sequence 14, Application US/09882529
: Patent No. US20020132317A1
: GENERAL INFORMATION:
: APPLICANT: Peyman, John A
: APPLICANT: da Silva, Antonio
: APPLICANT: Hockman, Paula
: TITLE OF INVENTION: NOVEL INTERFERON-INDUCED TETRASPAN PROTEIN AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING SAME
: FILE REFERENCE: 15966-771
: CURRENT APPLICATION NUMBER: US/09/882,529
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 60/721,565
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 14
: LENGTH: 116
: TYPE: prt
: ORGANISM: Homo sapiens
US-09-882-529-14

```

|    |  |        |                    |        |                                  |
|----|--|--------|--------------------|--------|----------------------------------|
|    | Query Match  | 4.0%;  | Score 179.5;       | DB 10; | Length 116;                      |
|    | Best Local Similarity  | 37.2%; | Pred. No. 8.8e+07; |        |                                  |
|    | Matches  | 45;    | Conservative       | 23;    | Mismatches 46; Indels 5; Gaps 2. |
| OY | 136 VNAANEDLLHGGLALALVYKAGFEIOESKOFVARYKVSAGELAVTGARLPCKQIIT       | 195    |                    |        |                                  |
|    | :    :    :    :    :    :    :    :    :    :    :    :           |        |                    |        |                                  |
| Dd | 1 VNAANSRLHGGVGAGVALARAKGKEAMPEA---FKKAPRCVPYGAEVLTLTGGGGLPAKYVIH  | 57     |                    |        |                                  |
| OY | 136 AVGPWMEMENDKOGCTGKLORATIVSLILNLYIKNTMIKTVAIPAALSSGIFQPFLNCTKTI | 255    |                    |        |                                  |
|    | :    :    :    :    :    :    :    :    :    :    :    :           |        |                    |        |                                  |
| Dd | 58 AVGNPNSKGCGEEEGDELLEKAYRAILR--LADENGNIKSVAFFLLSTGIIGGPKDRBAQGSL | 115    |                    |        |                                  |
| OY | 256 V 256  |        |                    |        |                                  |
| Dd | 116 L 116  |        |                    |        |                                  |

Search completed: May 13, 2003, 11:54:36  
Job time : 59 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:39:37 ; Search time 49 Seconds  
(without alignments)  
1675.486 Million cell updates/sec

Title: US-09-830-762-2  
Perfect score: 4447  
Sequence: 1 MDESMVAGAAAYNEKSGRIT.....PMRPFQHPWRGFGSGSPVD 854  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description               |
|------------|-------|-------------|--------|----|---------------------------|
| 1          | 291.5 | 6.6         | 599    | 2  | B72368 conserved hypothet |
| 2          | 288   | 6.5         | 177    | 2  | A99469 conserved hypothet |
| 3          | 235   | 5.3         | 176    | 2  | AF1419 hypothetical prote |
| 4          | 232   | 5.2         | 176    | 2  | AG1794 hypothetical prote |
| 5          | 220   | 4.9         | 173    | 2  | I39569 hypothetical prote |
| 6          | 211   | 4.7         | 173    | 2  | E83182 conserved hypothet |
| 7          | 204.5 | 4.6         | 186    | 2  | AD0636 conserved hypothet |
| 8          | 203   | 4.6         | 190    | 2  | E71027 hypothetical prote |
| 9          | 194   | 4.4         | 169    | 2  | T35937 hypothetical prote |
| 10         | 189   | 4.3         | 359    | 2  | T12540 hypothetical prote |
| 11         | 187.5 | 4.2         | 343    | 2  | H70537 probable lppd prot |
| 12         | 186.5 | 4.2         | 177    | 2  | G90806 probable polypote  |
| 13         | 186.5 | 4.2         | 177    | 2  | C85666 probable polypote  |
| 14         | 186.5 | 4.2         | 177    | 2  | B64847 probable polypote  |
| 15         | 185   | 4.2         | 186    | 2  | H75106 hypothetical prote |
| 16         | 183.5 | 4.1         | 657    | 2  | JC7785 TCD-Inducible pol  |
| 17         | 183   | 4.1         | 199    | 2  | H69439 conserved hypothet |
| 18         | 165.5 | 3.7         | 368    | 2  | I80811 histone H2A.1 - ra |
| 19         | 160   | 3.6         | 203    | 2  | T18653 conserved hypothet |
| 20         | 159.5 | 3.6         | 165    | 2  | E70385 conserved hypothet |
| 21         | 156   | 3.5         | 598    | 2  | T46327 hypothetical prote |
| 22         | 156   | 3.5         | 4085   | 2  | S28600 hypothetical prote |
| 23         | 155   | 3.5         | 1169   | 2  | A64505 p115 homolog - met |
| 24         | 151   | 3.4         | 194    | 2  | D72545 hypothetical prote |
| 25         | 150.5 | 3.4         | 191    | 2  | E84831 conserved hypothet |
| 26         | 148   | 3.3         | 170    | 2  | B75291 conserved hypothet |
| 27         | 147   | 3.3         | 716    | 2  | E89998 conserved hypothet |
| 28         | 147   | 3.3         | 2492   | 1  | C44213 nonstructural poly |
| 29         | 146   | 3.3         | 266    | 2  | G89797 conserved hypothet |

|    |       |     |      |   |                           |
|----|-------|-----|------|---|---------------------------|
| 30 | 146   | 3.3 | 1552 | 2 | G86344 T2211.2 protein -  |
| 31 | 145   | 3.3 | 728  | 2 | A81385 probable ATP/GTP   |
| 32 | 139   | 3.1 | 2492 | 1 | MNMVTD nonstructural poly |
| 33 | 138   | 3.1 | 2492 | 1 | A44213 nonstructural poly |
| 34 | 137   | 3.1 | 4488 | 1 | KRIHM2 genome polypote    |
| 35 | 136.5 | 3.1 | 689  | 2 | T40364 hypothetical prote |
| 36 | 136   | 3.1 | 2104 | 2 | T38774 myosin-3 heavy cha |
| 37 | 135   | 3.0 | 1899 | 2 | T32732 PAM C-terminal int |
| 38 | 135   | 3.0 | 1919 | 2 | T42098 synaptonemal compl |
| 39 | 133   | 3.0 | 845  | 2 | I48176 RESA-H3 antigen pr |
| 40 | 133   | 3.0 | 1558 | 2 | B71603 nucleotide-binding |
| 41 | 132.5 | 3.0 | 1620 | 2 | S61535 nonstructural poly |
| 42 | 132.5 | 3.0 | 2512 | 1 | MNMVS myosin heavy chain  |
| 43 | 132   | 3.0 | 1972 | 1 | A41604 chromosome segrega |
| 44 | 131.5 | 3.0 | 1233 | 2 | T30534 centromere protein |
| 45 | 131.5 | 3.0 | 2663 | 1 | S28261                    |

ALIGNMENTS

RESULT 1

B72368 conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72368  
R:Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,  
C.M.  
Nature 399, 323-329, 1999  
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72368  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-599 <ARN>  
A:Cross-references: GB:AE001727; GB:AE000512; NID:94981015; PIDN:AD35593.1; PID:9498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0508

Query Match 6.6%; Score 291.5; DB 2; Length 599;  
Best Local Similarity 31.9%; Pred. No. 1.8e-10;  
Matches 80; Conservative 47; Mismatches 91; Indels 33; Gaps 7;

|    |     |  |     |
|----|-----|--|-----|
| QY | 57  | GENYEMQIPININDE-----KILKNN-----EROLCEYLQKKFGCISTLVSPV        | 99  |
| DB | 359 | GGGYLY--PHDGGFVKTNVLYPEKLNKNEVIEQPKRVGEFEELFERLRKIM-----PE   | 408 |
| QY | 100 | QEGNSKSLQVFRKMLTPRIETLSVWRKDDLTTHAVDAVVAANEDLHGGGLALATVAKAGF | 159 |
| DB | 409 | KYGGSMAMEVREKLEKKKIRIVKCDITREVDIVVANAENYLNKGGVAGAIYRAGGS     | 468 |
| QY | 160 | EIQESKQFVARVYKVSAGEIAVTGAGRLPCQIIHAVGPRMWDKQCTGKIORAIVS      | 219 |
| DB | 469 | VIOESDRIIVDERGRVPTGEAVVTSAGKLAKAYVHTGVPW----RGSGHDEDLTYKA    | 524 |
| QY | 220 | ILNVIYKN-THIKTVAIIPALSSGTFOPPLNCTTIYETITRVSLOGPMNSNKEIHLV    | 278 |
| DB | 525 | VNALRAHEKLKSTISMAISTGIFGPKERAVGIFSKAIDFDIHP-DITTEIRIC        | 583 |
| QY | 279 | SNEDPTVAAFK 289  |     |
| DB | 584 | NIDETTKIFE 594   |     |

RESULT 2  
A99469 conserved hypothetical protein [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
C:Accession: A99469





Db 116 RLAFNISTGTGTAFFPRERADIAIAAVALAALAAAP--EIEQVTFVCEDDENTRLXR 169

# RESULT 6

conserved hypothetical protein PA3693 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 23-Mar-2001

C:Accession: E83182

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Watterer, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lam,

.Loay, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: E83182

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-173 <STO>

A:Cross-references: GB:AE004789; GB:AE004091; NID:g9949857; PIDN:AAG07081.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3693

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.7%; Score 211; DB 2; Length 173;

Best Local Similarity 35.2%; Pred. No. 3.1e-06;

Matches 57; Conservative 28; Mismatches 65; Indels 12; Gaps 5;

QY 119 ELSTVMDLITTHAVDAVNAANEDLHGGGLALVAKAGFEIOESKQFVARVYKVSAG 178

Db 3 EVRWOGDITRLAVDAIVNAANSLGGGVDGDAIHRAAGAEIVACRL--HG-CKTG 58

QY 179 ELAVTGAGRLPCKQIIHAGPRMMEWDKQCTGKIQ-RAIVSILNVIVYKNTIKTVAIP 237

Db 59 EAKITGFPLPAHVHTVGPW---RCGDNGEAEALLASCYRSLALAEQACAAVAF 114

QY 238 AASSGIFPPLNCTKTIVETIRVSLQGRPMNSNKEIHVS 279

Db 115 AISCIGYGPLRQAAIAVEV---CRORPAHSLEEIVLVA 153

# RESULT 7

AD0636

conserved hypothetical protein STY1184 [imported] - Salmonella enterica subsp. enterica

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AD0636

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; PMID:11677608

A:Accession: AD0636

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-186 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08271.1; PID:g16502318; GSPDB:GN00176

C:Genetics:

A:Gene: STY1184

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.6%; Score 204.5; DB 2; Length 186;

Best Local Similarity 32.7%; Pred. No. 8.9e-06;

Matches 48; Conservative 33; Mismatches 61; Indels 5; Gaps 3;

QY 113 MLTPRIELSVKDDLTTHAVDAVNAANEDLHGGGLALVAKAGFEIOESKQFVARV 172

Db 7 VMHSR--LOVIGDITQLSVDAIVNAANASLMGGGVDGDAIHRAAGPALDLACKLROO 64

QY 173 GKVSAGEIAVAGRLPCKQIIHAGPRMMEWDKQCTGKIQRAIVSILNVYKNTIK 232

Db 65 GEGQGTCHAVITPAGKLSAAVAIVHTVGPVWRGGEQDAE-LLEEAVNCL--LLAEANHR 121

QY 233 TVAIPALSSGIFPPLNCTKTIVETI 259

Db 122 SIAFPATSTGVGYPRQAQAEVAVRTV 148

# RESULT 8

E71027

hypothetical protein PH1513 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000

C:Accession: E71027

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohikubo, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: E71027

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-190 <KAM>

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BA30621.1; PID:g3257938

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenB

C:Genetics:

A:Gene: PH1513

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.6%; Score 203; DB 2; Length 190;

Best Local Similarity 34.2%; Pred. No. 1.1e-05;

Matches 66; Conservative 38; Mismatches 59; Indels 30; Gaps 12;

QY 124 KDDLTHAVDAVNAANEDLHGGGLALVAKAGFEIOE---SKQFVAR-YGK-VSA 177

Db 9 RGDITFRFAEIVNAANKYLEHGGGVAIVAAASGVSEYTRISKEEMRROLGKDMIEH 68

QY 178 GEIAVAGARLP---CKQIIHAGPR---MMEMDKQCTGKIQRAIVSILNVYKNTIK 230

Db 69 GEVYVTPPMKLEKNGKVIYHTVGPYCGGVSK-DKE---EKIKAILALAKKA--DEIG 122

QY 231 IKTVAPALSSGIFPPLNCTKTIVETIRVSLQGRPMNSNKEIHV--SNEDPTVAAF 288

Db 123 VKSIAPPAISAGICGPLEKVEVTFREYVKEFLK--VANHVKVYLVLYSEED-----Y 174

QY 289 KAASEPT-LGKSE 300

Db 175 KLALETIGLDND 187

# RESULT 9

T35937

hypothetical protein SC9B5.17 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 23-Mar-2001

C:Accession: T35937

R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M

submitted to the EMBL Data Library, January 1999

A:Reference number: Z21551

A:Accession: T35937

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-169 <SEB>

A:Cross-references: EMBL:AL035206; PIDN:CAA22759.1; GSPDB:GN00070; SCOEDB:SC9B5.17

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC9B5.17

C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.4%; Score 194; DB 2; Length 169;

Best Local Similarity 30.5%; Pred. No. 3.5e-05;

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: H70517  
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-343 <COL>  
A:Cross-references: GB:Z97193; GB:AL123456; NID:g3261816; PIDN:CAB10035.1; PID:e32487  
C:Genetics:  
A:Experimental source: strain H37RV  
C:Gene: lppD

Query Match 4.2%; Score 187.5; DB 2; Length 343;  
Best Local Similarity 29.0%; Pred. No. 0.00026;  
Matches 56; Conservative 32; Mismatches 74; Indels 31; Gaps 5;

OY 118 TELSWKDDITTHAVDAVVNAANEDLLHGGGLATLVAGGFEIOESKQCVARIGKSA 177  
| | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 174 IELEHVQDVTKLELDATITNAANTRLRHAGVAAAIARAGPELORESTE---KAPIGL 229  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 178 GEIAVTGARLPCKOIHAV-----GPRMMEMDKCGCTGKLORAVISILNVIIYNTHIK 232  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 230 GEAVETTAGDMPARVIHAAMTELGPSTSGELITTAATATLRKA-----DELGR 279  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 233 TVAIRPALSSTGFOPPLNLCTTIYETTIVTSIQGPKMSNLKEIHLVSNEDEPTVAEFKAAS 292  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 280 SLATAVFQTGGFPLDPAARLMGVARRHRPG-----SLRVVFVAVHGDAAEAFSA-- 333  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 293 EFILGSELGOET 305  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 334 -----IQAGEDT 340  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 12  
g90806  
Probable polypeptide [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: G90806  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hatvoti, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: G90806  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-177 <NAV>  
A:Cross-references: GB:BA000007; PIDN:BAB34846.1; PID:g13360887; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: Ecs1423  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.2%; Score 186.5; DB 2; Length 177;  
Best Local Similarity 32.2%; Pred. No. 0.00011;  
Matches 46; Conservative 29; Mismatches 65; Indels 3; Gaps 2;

OY 117 RIELSVMKDDITTHAVDAVVNAANEDLLHGGGLATLVAGGFEIOESKQCFVARIGKVS 176  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 2 KTRIHHVOGDITTKLAVDIVNAANPSLMGGCGVGGAITHRAAGPALDLACLVRROOGDCP 61  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 177 AGEIAVTCAGRRLPCKOITIHAGPRMMEMDKGCTGKLORALVSIILNVIIYNTHIKTVAI 236  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 62 TGHAVITACGLPAKAVVHTGTGFW--RGGEONEDOLLADATLNSLRLVANS--YTSAVF 118  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
OY 237 PALSSGIFOPPLNLCTKIIVETI 259  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
DB 119 PAISTGVGYDPRAAAAEIATVKTV 141  
| | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 13  
c85666  
Probable polypeptide Z1679 [imported] - Escherichia coli (strain O157:H7, substrain E

C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: C85666  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Hiller, U.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanata, E.; Polmanousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85666  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-177 <STO>  
A:Cross-references: GB:AE005174; NID:912514577; PIDN:AAG55791.1; GSPDB:GN00145; UWGP:216  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: 21679  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.2%; Score 186.5; DB 2; Length 177;  
Best Local Similarity 32.2%; Pred. No. 0.00011;  
Matches 46; Conservative 29; Mismatches 65; Indels 3; Gaps 2;

OY 117 RLEIYVWMDLTHAVDAVNNANEDLLHGGGLALALYKAGGFEOESKQFVARVKYS 176  
DB 2 KRIHVVGDDITKLAVDIYVNNANPSLMGGGVGDAIHRAPALDACLKVRQOQDCP 61  
OY 177 AGEIAVGTAGRLPCQKQIIVHAGPRMMEMDKQCTGKLORAIVSIINVIYKNTHTIKTVAI 236  
DB 62 TGHAVITLAGDLPAAKAVHTVGPVW-RGGEONEDDQLDQAYLNSLRVAANS--YTSVAF 118  
OY 237 PALSGIFQFPLNCTKTIVETI 259  
DB 119 PAISTGVYGYPRAAAEIAVKTV 141

## RESULT 14

B64847  
Probable polypeptide b1045 [similarity] - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: B64847  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: B64847  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-177 <BLAT>  
A:Cross-references: GB:AE000206; GB:U00096; NID:91787282; PIDN:AAC74129.1; PID:91787283;  
A:Experimental source: strain K-12, substrain MG1655  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.2%; Score 186.5; DB 2; Length 177;  
Best Local Similarity 32.2%; Pred. No. 0.00011;  
Matches 46; Conservative 29; Mismatches 65; Indels 3; Gaps 2;

OY 117 RLEIYVWMDLTHAVDAVNNANEDLLHGGGLALALYKAGGFEOESKQFVARVKYS 176  
DB 2 KRIHVVGDDITKLAVDIYVNNANPSLMGGGVGDAIHRAPALDACLKVRQOQDCP 61  
OY 177 AGEIAVGTAGRLPCQKQIIVHAGPRMMEMDKQCTGKLORAIVSIINVIYKNTHTIKTVAI 236  
DB 62 TGHAVITLAGDLPAAKAVHTVGPVW-RGGEONEDDQLDQAYLNSLRVAANS--YTSVAF 118  
OY 237 PALSGIFQFPLNCTKTIVETI 259  
DB 119 PAISTGVYGYPRAAAEIAVKTV 141

## RESULT 15

H75106

hypothetical protein PAB0445 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: H75106  
R:Anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s  
A:Reference number: A75001  
A:Accession: H75106  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-186 <KAW>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:95458067; PIDN:CA849569.1; PID:9545  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0445  
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0445

Query Match 4.2%; Score 185; DB 2; Length 186;  
Best Local Similarity 32.2%; Pred. No. 0.00015;  
Matches 57; Conservative 33; Mismatches 61; Indels 26; Gaps 9;

OY 126 DITTHAVDAVNNANEDLLHGGGLALALYKAGGFEL-----QESKQFVARVKYSAG 178  
DB 11 DITTRKAEIAVNAANKYLEHGGGVAVAIKAAASGVSEYIRISKEMRKQIGR-DWIEHG 69  
OY 179 EIAVT---GAGRUPCQKQIIVHAGP---RMEMDKQCTGKLORAIVSIINVIYKNTHTI 231  
DB 70 EVAVITPPLANLNGKYYVHTVGPICGKNDE-DKR--KLEIALGLAKKA--DELGV 123  
OY 232 KTVAIIPALSSGIFQFPLNCTKTIVETIIVSLQGRPMNSLKEIHLV--SNEDPTVA 286  
DB 124 RSIAPFAISAGIYGCPLFEVVKTFKLVNVEFLKS--AKNVVDVYLVLYSERDYEA 177

Search completed: May 13, 2003, 11:45:29  
Job time : 54 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:30:37 ; Search time 27 Seconds

(Without alignments)  
1311.881 Million cell updates/sec

Title: US-09-830-762-2

Perfect score: 4447  
Sequence: 1 MDFSVMGAAAYNEKSGRIT.....PMRPFAPWRCFASGSPVD 854

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID         | Description         |
|------------|-------|-------------|--------|------------|---------------------|
| 1          | 291.5 | 6.6         | 599    | Y508_THEMA | O9wyx8 thermotoga   |
| 2          | 291   | 6.5         | 182    | YN83_SUITO | O96xy5 sulfolobus   |
| 3          | 288   | 6.5         | 177    | YS99_SULSO | O97uud sulfolobus   |
| 4          | 267.5 | 6.0         | 182    | YB11_PYRAE | O8zxc3 pyrobaculum  |
| 5          | 235   | 5.3         | 176    | YR59_LISMO | O93s63 listeria mo  |
| 6          | 232   | 5.2         | 176    | YT02_LISIN | O926v8 listeria in  |
| 7          | 228   | 5.1         | 183    | Y189_ACISP | O93sxt actinobact   |
| 8          | 220   | 4.9         | 173    | YGB2_ALCEU | O440z0 alcaligenes  |
| 9          | 211   | 4.7         | 173    | Y0J3_PSEAE | O3hxuv pseudomonas  |
| 10         | 203.5 | 4.6         | 179    | YMDB_SALTY | O8zqz9 salmonella   |
| 11         | 203   | 4.6         | 190    | YF13_PVRHO | O59182 pyrococcus   |
| 12         | 201.5 | 4.5         | 171    | Y334_RALSO | O8y2k1 ralistonia s |
| 13         | 194   | 4.4         | 169    | YSP0_STRCO | O9zbt3 streptomyce  |
| 14         | 193   | 4.3         | 181    | Y189_STRNO | O9ey1c streptomyce  |
| 15         | 187.5 | 4.2         | 359    | Y199_MCTU  | O07733 mycobacteri  |
| 16         | 186.5 | 4.2         | 177    | YMDB_ECOLI | P75918 escherichia  |
| 17         | 185   | 4.2         | 186    | Y445_PYRAB | O9v0y3 pyrococcus   |
| 18         | 184   | 4.1         | 177    | Y189_STRGR | O9khez streptomyce  |
| 19         | 183   | 4.1         | 192    | YF21_ARCFU | O88751 archaeoglob  |
| 20         | 182.5 | 4.1         | 176    | Y72X_RHTLO | O985d2 rhizobium l  |
| 21         | 171   | 3.8         | 196    | YB05_THEAC | O9hj6f thermoplasma |
| 22         | 168   | 3.8         | 186    | Y189_THREO | O97au0 thermoplasma |
| 23         | 166   | 3.7         | 261    | Y189_TREMD | O93rg0 treponema m  |
| 24         | 162.5 | 3.7         | 371    | H2AY_HUMAN | O7536f homo sapien  |
| 25         | 159.5 | 3.6         | 165    | Y987_AQUAE | O67112 aquifex aeo  |
| 26         | 157   | 3.5         | 370    | H2AY_RAT   | O02874 ratulus notv |
| 27         | 156   | 3.5         | 4085   | RRPA_CVH22 | O05002 human, coron |
| 28         | 153   | 3.4         | 1169   | SMC_METJA  | O59037 methanococc  |
| 29         | 151   | 3.4         | 194    | YG48_AERPE | O9ybe3 aeropyrum p  |
| 30         | 148   | 3.3         | 170    | YMB8_DEIRA | O9ts39 deinococcus  |
| 31         | 147.5 | 3.3         | 2485   | POLN_EEYV3 | P36327 venezuelan   |
| 32         | 146   | 3.3         | 266    | Y325_STAMM | O99wq1 staphylococ  |
| 33         | 139   | 3.1         | 2492   | POLN_EEYV1 | P27282 venezuelan   |

|    |       |     |      |            |                     |
|----|-------|-----|------|------------|---------------------|
| 34 | 138   | 3.1 | 2492 | POLN_EEYV1 | P36328 venezuelan   |
| 35 | 137   | 3.1 | 1166 | TNK2_HUMAN | O9hk22 homo sapien  |
| 36 | 137   | 3.1 | 4488 | RRPA_CVH22 | P19781 murine coro  |
| 37 | 136   | 3.1 | 2104 | MTS3_SCHPO | O14157 schizosacch  |
| 38 | 135   | 3.0 | 1919 | HAPP_RAT   | P97924 ratulus norv |
| 39 | 133   | 3.0 | 845  | SCPI_MESAU | O60563 mesocricetu  |
| 40 | 132.5 | 3.0 | 1012 | XC05_HUMAN | O75665 homo sapien  |
| 41 | 132.5 | 3.0 | 2512 | POLN_SINDV | P03317 sindbis vir  |
| 42 | 132   | 3.0 | 904  | SYA_SULFO  | O97134 sulfolobus   |
| 43 | 132   | 3.0 | 1327 | TNKL_HUMAN | O95271 homo sapien  |
| 44 | 132   | 3.0 | 1972 | MYHB_HUMAN | P35749 homo sapien  |
| 45 | 132   | 3.0 | 1972 | MYHB_RABIT | P35748 oryctolagus  |

## ALIGNMENTS

```

RESULT 1
Y508_THEMA STANDARD; PRT; 599 AA.
ID Y508_THEMA
AC O9wyx8:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein TM0508.
GN TM0508.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=MSRB / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Ginn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
-i - SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE UPF0189
FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AE001727; AAD55593.1; -
CC TIGR: TM0508; -
CC InterPro: IPR002589; A1PP.
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR003959; AAA_ATPase-centr.
CC InterPro: IPR000862; RFGdomain.
CC Pfam: PF00004; AAA; 1.
CC Pfam: PF01661; A1PP; 1.
CC SMART: SM00506; A1PP; 1.
CC SMART: SM00382; AAA; 1.
CC Hypothetical protein: ATP-binding, Complete proteome.
CC NP_BIND 49 56 ATP (POTENTIAL).
CC DOMAIN 422 599 UPF0189.
CC SEQUENCE 599 AA; 67417 MW; BOFBSCB87155BC7 CRC64;
Query Match 6.6%; Score 291.5; DB 1; Length 599;
Best Local Similarity 31.9%; Pred. No. 9.3e-11;
Matches 80; Conservative 47; Mismatches 91; Indels 33; Gaps 7;

```

QY 57 GENYSWOPIPIINHDF-----KILKNN-----EROLCEVLONKEGCIJSTLVSFV 99  
 DB 359 GEGYLY--PHDFGFGFVKTNYLPEKLNKEVIFQPKRGVEEELERLKLW-----PE 408  
 QY 100 OEGSKSLQVFRKMLTPRIELSVKMDLTHAVDAVYVNAANEDLLHGGGLALAYKAGCF 159  
 DB 409 KYGGSMAEYKLELYGKRTIRYKGTITREEDAVYVNAANEYKHGGGVALVRAGGS 468  
 QY 160 EIOESKOFVARYKVSAGELAVTAGRLPCQITIHAVGPRMWMEDKOGCTGLQRAIVS 219  
 DB 469 VIOESDRIVQERGRVPTGEAVVTSAGKLAKKYIHTVGPW-----RGSGDEDELLYKA 524  
 QY 220 ILAVVITK--THITVAIPALSSGIFQFPPLNCTKTIVETIRVSLQCKPMKSNKEIHLV 278  
 DB 525 VYNALLLAHELEKLSISMIPALSTGIFGPKERAVGIFSKAIRFDIDHP--DTLLEIRIC 583  
 QY 279 SNEDEPTVAAFK 289  
 DB 584 NIDEETKIFE 594

RESULT 2  
 YN83\_SULTO STANDARD; PRT; 182 AA.  
 AC 096XY5:  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ST2383.  
 GN ST2383.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=111955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JCM 10545 / 7;  
 RX MEDLINE=21456156; PubMed=11572479;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Anai A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Oshima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
 RL DNA Res. 8:123-140(2001).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AP000989; BAB67492.1;  
 DR InterPro: IPR002589; Alpp. 1.  
 DR Pfam: PF01661; Alpp. 1.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 182 AA; 20183 MW; 1E0A072586AF8947 CRC64;  
 Query Match 6.5%; Score 291; DB 1; Length 182;  
 Best local Similarity 38.5%; Pred. No. 1.9e-11;  
 Matches 62; Conservative 37; Mismatches 46; Indels 16; Gaps 4;  
 QY 120 LSVKMDLTHAVDAVYVNAANEDLLHGGGLALAYKAGCFEIOESKOFVARYKVSAGE 179  
 DB 3 VKIKGITELEAEAVIYVNAANSYLEHGGVAVRIVEKGYTIQESREYVKKYGPVPTG 62  
 QY 180 IAVTAGRLPCQITIHAVGPRMWMEDKOGCTG--KLORAIVSYLVYVYKNTHTKVAIP 237  
 DB 111 IAVTAGRLPCQITIHAVGPRMWMEDKOGCTG--KLORAIVSYLVYVYKNTHTKVAIP 237

DB 63 VAVTSAGKLAKYVIAHAVGPRY-----GIEGEEKLEAEAIRNALRKA--BELKLSSIALP 114  
 QY 238 ALSSGIFQFPPLNCTKTIVETIRVSLQCKPMKSNKEIHLV 278  
 DB 115 AISTGITYGYPEICAEKMKVYI-----KEEYTFEKLHNTI 149

RESULT 3  
 YS99\_SULSO STANDARD; PRT; 177 AA.  
 AC 097JUN4:  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein SSO2899.  
 GN SSO2899.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxID=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweyer M.J., Chan-Welner C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moers A., Erauso G., Fletcher C., Gordon P.M.K.,  
 RA Helkamp-de Jong I., Jaffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Thériault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gasterjerud T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,  
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AE006882; AAK43008.1;  
 DR InterPro: IPR002589; Alpp.  
 DR Pfam: PF01661; Alpp. 1.  
 DR SMART: SM00506; Alpp. 1.  
 KW Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 177 AA; 19445 MW; 8353685B14FEC9A96 CRC64;

Query Match 6.5%; Score 288; DB 1; Length 177;  
 Best local Similarity 40.7%; Pred. No. 2.7e-11;  
 Matches 59; Conservative 33; Mismatches 43; Indels 10; Gaps 3;  
 QY 118 IELSVKMDLTHAVDAVYVNAANEDLLHGGGLALAYKAGCFEIOESKOFVARYKVSAGE 177  
 DB 8 LEYLINGDITTELEADAVIYVNAANSYLDHGGVAVRIVEKGYTIQESREYVKKYGPVPTG 67  
 QY 178 GEIIVTAGRLPCQITIHAVGPRMWMEDKOGCTG--KLORAIVSYLVYVYKNTHTKVA 235  
 DB 68 GEVAVVTSAGKLAKYVIAHAVGPRY-----GIEGEEKLEAEAIRFSL--LAADELSSSIA 119  
 QY 236 IPALSSGIFQFPPLNCTKTIVETIR 260  
 DB 120 MPALSTGITYGYPEICARIMANVYLK 144  
 RESULT 4  
 YB11\_PYRAE STANDARD; PRT; 182 AA.  
 ID YB11\_PYRAE  
 AC 082XT3;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein PAE111.
GN PAE111.
OS Pyrobaculum aerophilum
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
ON NCBI_TaxID=13773;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum".
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE009804; AAL63263.1;
DR InterPro: IPR002589; A1PP.
DR Pfam: PF01661; A1PP: 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 182 AA; 19369 MW; 3158ADD9C2A3EF3E CRC64;

Query Match
Best Local Similarity 6.0%; Score 267.5; DB 1; Length 182;
Matches 64; Conservative 33; Mismatches 44; Indels 31; Gaps 5;

QY 118 IELSVWKDDLTTHAVDAVYNNANEDLHGGLALALVAGGEIOEESKQFVARYKVS 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 8 VEVVLMRGDITREVDADVNAANSYLEHSGVAGVIRKGVIOEESREVRKRKGPVPV 67

QY 178 GEIAVTGAGRLPCKOIHAAGPRMWMEMDKGCGTGLQRAIVSILNYVYKNTHTK 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 68 GDVAVTSAGRIKAKYVHAAGPR-----CGVEPIKLAEAV-----KNALIKA 110

QY 233 -----TVAIPALSSGIFQFPLNCTKTIVETIRVSLGCKPMNSLKEIHV 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 111 EELGVSTALPAISGIRGCPYDAAEQMAVIR---EVAPALRSIRILV 159

RESULT 5
YR59_LISMO
ID YR59_LISMO STANDARD; PRT; 176 AA.
AC 08Y353;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Lmo2759.
GN LMO2759.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
ON NCBI_TaxID=1639;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RC MEDLINE=21537279; PubMed=11679669;
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouani F., Couve E., de Darvar P., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entlan K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaeerst U., Kreft J., Kuhn M., Kunst F., Kutnapkat G.,
RA Jones L.-M., Kaeerst U., Kreft J., Kuhn M., Kunst F., Kutnapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,

```

```

RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlutner T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.".
RL Science 294:849-852(2001).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL591984; CAPD00972.1;
DR ListList; LMO02759;
DR InterPro: IPR002589; A1PP.
DR Pfam: PF01661; A1PP: 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 176 AA; 18984 MW; 3362EA4DA723050B CRC64;

Query Match
Best Local Similarity 5.3%; Score 235; DB 1; Length 176;
Matches 55; Conservative 39; Mismatches 73; Indels 4; Gaps 3;

QY 118 IELSVWKDDLTTHAVDAVYNNANEDLHGGLALALVAGGEIOEESKQFVARYKVS 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MEITIVKGDITREVDADVNAANSGLGGVDGAIHQAAGPDLKECEVYINRKGSP 60

QY 178 GEIAVTGAGRLPCKOIHAAGPRMWMEMDKGCGTGLQRAIVSILNYVYKNTHTK 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 GEIAVTSAGDKAKYVHAAGPIWKGDEHGE-ANKLASCYMKALDLAAGKD--LTSIAFP 117

QY 238 ALSSGIFQFPLNCTKTIVETIRVSLGCKPMNSLKEIHVSNEDPTAAAF 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 118 NISTGVGFPKLAEVALYVRKAAE-EYDTSIKELRFVCEDEENLKL 167

RESULT 6
YR02_LISIN
ID YR02_LISIN STANDARD; PRT; 176 AA.
AC 09Z618;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Lin2902.
GN Lin2902.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
ON NCBI_TaxID=1642;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RC MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouani F., Couve E., de Darvar P., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entlan K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaeerst U., Kreft J., Kuhn M., Kunst F., Kutnapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlutner T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.".
RL Science 294:849-852(2001).
CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```





CC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 RX MEDLINE=20437337; PubMed=10964043;  
 RA Slover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Ketzler J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE004789; AAG07081.1; -  
 DR InterPro: IPR002589; AIPP.  
 DR Pfam: PF01661; AIPP.1.  
 DR SMART: SM00506; AIPP.1.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 173 AA; 18219 MW; D94A49C8FD208BF8 CRC64;  
 SQ  
 Query Match 4.7%; Score 211; DB 1; Length 173;  
 Best Local Similarity 35.2%; Pred. No. 1.6e-06;  
 Matches 57; Conservative 28; Mismatches 65; Indels 12; Gaps 5;  
 OY 119 ELVWKDULTHAVDAVYVNAANEDLLHGGGLALVKAAGFEIOESKOFVARYKVSAG 178  
 Db 3 EVRWGGITRIADVAIYNAANSSLLGGGVDGAIHRAAGAEIVVACRL--HG-CKTG 58  
 OY 179 ELVATGAGRLPCKOIITHAVGPRMMEWDKQCTGKLO-RAIYSILNVYKTKHTKTVAP 237  
 Db 59 EAKITRGRLPAHYIHTVGVW----RGDNGEAEILLASCTRRSLALAEAGASVAP 114  
 OY 238 ALSSGIFQPLNCTKTIYETIRVSLQCKPMMSNLKEIHLVS 279  
 Db 115 AISCIGYGPLEQAAIAVEEY---CRQRPAAHSSLEIIVVA 153  
 RESULT 10  
 YMDB\_SALTY STANDARD: PRT; 179 AA.  
 ID YMDB\_SALTY 082029; 0827M1;  
 AC 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Hypothetical protein ymdb.  
 GN YMDB OR STM1147 OR STY1184.  
 OS *Salmonella typhimurium*, and  
 CC *Salmonella typhi*.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC *Salmonella*.  
 OX NCBI\_TaxID=602, 601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium

RT LT2.";  
 RL Nature 413:852-856(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.F.G., Sebaiha M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrall B.G.;  
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE008749; AAL20077.1; -  
 DR EMBL: AL627269; CAB08271.1; ALT\_INIT.  
 DR StyGene; SG77777; ymdb.  
 DR InterPro: IPR002589; AIPP.  
 DR Pfam: PF01661; AIPP.1.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 179 AA; 19196 MW; 05AF68EF0CD9315 CRC64;  
 SQ  
 Query Match 4.6%; Score 203.5; DB 1; Length 179;  
 Best Local Similarity 32.9%; Pred. No. 4.8e-06;  
 Matches 46; Conservative 31; Mismatches 60; Indels 3; Gaps 2;  
 OY 120 LSWKDDULTHAVDAVYVNAANEDLLHGGGLALVKAAGFEIOESKOFVARYKVSAGE 179  
 Db 5 LQVTDGDTTQSVDAIYNAANSSLLGGGVDGAIHRAAGAPLDDCKLIRQDQEGCQTGH 64  
 OY 180 IAVTCAGRLPCKOIITHAVGPRMMEWDKQCTGKLORAIYSILNVYKTKHTKTVAPAL 239  
 Db 65 AVTPACKLSAKAVIHTVGPVWRGGEHQEAE-LLEEAVRNL--LLAANPFRSIAFPAPAI 121  
 OY 240 SSGIFQPLNCTKTIYETI 259  
 Db 122 STGVYGPRAQAAEVAVRTV 141  
 RESULT 11  
 YF13\_PYRHO STANDARD: PRT; 190 AA.  
 ID YF13\_PYRHO 059182;  
 AC 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Hypothetical protein PH1513.  
 GN PH1513.  
 OS *Pyrococcus horikoshii*.  
 CC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 CC *Pyrococcus*.  
 OX NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kanarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,

```

RA Aochi K.I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Maschi K.Y., Shizuya H., Kikuchi H.:
RT "Complete sequence and gene organization of the genome of a hyper-
RL thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC
CC -1- SIMILARITY: BELONGS TO THE DPF0189 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP000006; BAA30621.1;
DR InterPro: IPR002589; Alpp.
DR Pfam: PF01661; Alpp; 1.
DR SMART: SM00506; Alpp; 1.
KW Hypothetical protein; Complete proteome.
KW SEQUENCE 190 AA; 21106 MW; B5DEB0761BE3A684 CRC64;
SQ
Query Match 4.6%; Score 203; DB 1; Length 190;
Best Local Similarity 34.2%; Pred. No. 5,6e-06;
Matches 66; Conservative 38; Mismatches 59; Indels 30; Gaps 12;
QY 124 KDDLTHAVDAVVAANEDLLHGGGLALALYKAGGFELQEE--SKGFVAR-YGK-VSA 177
DB 9 RGDITKFAFAEIVVNAANKYLEHGGCVAAIAAKAASGDSEYTRISKEMRQLGKDWIEH 68
QY 178 GEIVAVTGGRLP---CKQIIHAVGPR---MMENDKQCGCTKLORAIYSILNYIYKNTH 230
DB 69 GEVVATTPMKLKEKGVKVIHTVPGYCGGVSK--DKE--EKLKAILCALKKA--DELG 122
QY 221 IKTAIPAALSSGFIQFPINLCTKTIVETIRYSLQGGPKMSMLKELTHLV--SNEDPTVAAF 288
DB 123 VKSIAPFALISAGITGCPLEKVEYRFRFEKVEVKEFLK---VANNVKEVYLVLYSERD----Y 174
QY 289 KAASEFT-LCKSE 300
DB 175 KLALETITGLGDND 187
RESULT 12
Y334_RALSO STANDARD; PRT: 171 AA.
AC 0812K1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein RSC0334.
GN RSC0334 OR RS03301.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
CC
OX NCBI_Taxid=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11000;
RC MEDLINE=21681879, PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brotlier P., Camus J.C., Cattolico L.,
RA Chandier M., Cholsne N., Claudel-Renard C., Gunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigüer P., Thebaud P., Whalen M., Winkler P., Levy M.,
RA Weissensbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
CC
CC -1- SIMILARITY: BELONGS TO THE DPF0189 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR EMBL; AL646058; CAD13862.1; -
DR InterPro; IPR002589; A1PP. Pfam; PF01661; A1PP; 1.
KW Hypothetical protein; Complete proteome.
SEQUENCE 171 AA; 18018 MW; 7DC99AF217882657 CRC64;

Query Match      4.5%; Score 201.5; DB 1; Length 171;
Best Local Similarity 34.2%; Pred. No. 6e-06;
Matches 52; Conservative % 24; Mismatches 61; Indels 15; Gaps 4;

QY 113 MLTPRIELSVKDDLTTHAVDAVVAANEDLLHGGLALVLVKGGCFIEESKFVARY 172
   | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MPPIPVTLRALRADITTLACDPAIVNAANSALGSGGVGAHRAAGPFLERCR---ALH 57
   | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 173 GKVSAGEIAYTAGRHLPCKQIIHAVGPWMEMDKQGCTGLORAIYSILNY---VIYKN 228
   | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 G-CRTGQAQAKINPGFLLPARYYIHFTVGPIWR-----CGRDEAALLAACYNRSLSLAKQ 109
   | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 229 THIKTVAILPALSSGIFFPFPLNCTTYETIR 260
   ::::: | | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 110 HDVRTIAFPICISTGYVGFPQLAPIAVRTVR 141
   ::::: | | | | | : | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
YSFO_STRCO          STANDARD;             PRT;    169 AA.
AC Q9ZBG3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein SC06450.
GN SC06450 OR SC985.17
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID:1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145.
RX MEDLINE=21996410; Pubmed=12000953;
RA Thomson S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Benson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kissler T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter K.,
RA Seeger K., Saunders D., Sharp S., Squares S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC The European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL035206; CAAT2759.1; -.
DR InterPro; IPR002589; A1PP.
DR Pfam; PF01661; A1PP; 1.
DR SMART; SM00506; A1PP; 1.
KW Hypothetical protein; Complete proteome.
SEQUENCE 169 AA; 17767 MW; 86965DB33C6E4740 CRC64;

```

Query Match 4.4%; Score 194; DB 1; Length 169;  
 Best Local Similarity 30.5%; Pred. No. 1.7e-05;  
 Matches 53; Conservative 35; Mismatches 72; Indels 14; Gaps 5;

QY 120 LSWMKDDLTTHAVDAVYVNAANEDLLHGGLALALVKAGFEIOESKQVARY-OK-VSA 177  
 DB 4 ITLVGGDITRQADALVNAANSLLGGGGVGAHRRGGPAILAECCRLRAGHLGGLPT 63

QY 178 GEIAVYAGRLPCKOIITHAVGPRMWM-DKQCGTGLORAIYSILNYVYKMTHTKTVAI 236  
 DB 64 GRAVATTAAGDLDARWYIHVGPVWSTEDRSGLLASCYRESLRAD-ELGARTVAF 118

QY 237 PALSSGIFQPLNCTKTIVETIRYSLOQKPMNSUKETHLVSNEDPVAAFKA 290  
 DB 119 PAISTGVYRMPMDARIAIVETVAT-KTSTVEIRFVLFDARAYAEFAA 166

RESULT 14  
 Y189\_STRNO STANDARD; PRT; 181 AA.  
 AC Q9EXY6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 19.0 kDa protein in sno 5' region (ORF7).  
 OS Streptomyces nogalater.  
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=38314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 27451;  
 RX MEDLINE=21539456; PubMed=11683270;  
 RA Torkell S., Kunari T., Palmu K., Meentsemae P., Hakala J.,  
 RA Ylilomko K.;  
 RA "The entire nogalamycin biosynthetic gene cluster of Streptomyces  
 RT nogalater: characterization of a 20-kb DNA region and generation of  
 RT hybrid structures.";  
 RL Mol. Genet. Genomics 266:276-288(2001).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0189 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF323753; AAG42849.1; -;  
 DR InterPro: IPR002589; Alpp.  
 DR Pfam: PF01661; Alpp; 1.  
 DR SMART: SM00506; Alpp; 1.  
 DR Hypothetical protein.  
 KW Hypothetical protein.  
 SQ SEQUENCE 181 AA; 19048 MW; C91684D0A793F4CA CRC64;

Query Match 4.3%; Score 193; DB 1; Length 181;  
 Best Local Similarity 32.9%; Pred. No. 2.2e-05;  
 Matches 48; Conservative 28; Mismatches 62; Indels 8; Gaps 4;

QY 120 LSWMKDDLTTHAVDAVYVNAANEDLLHGGLALALVKAGFEIOESKQF-VARYK-VSA 177  
 DB 4 ITLVGGDITRQADALVNAANSLLGGGGVGAHRRGGPAILAECCRLRAGHLGGLPT 63

QY 178 GEIAVYAGRLPCKOIITHAVGPRMWM-DKQCGTGLORAIYSILNYVYKMTHTKTVAI 236  
 DB 64 GRAVATTAAGDLDARWYIHVGPVWSTEDRSGLLASCYRPSRLAG-ELGARTVAF 118

QY 237 PALSSGIFQPLNCTKTIVETIRVS 262  
 DB 119 PALSTGVYRMPMDARIAIVETVRTT 144

RESULT 15  
 Y199\_MYCTU STANDARD; PRT; 359 AA.  
 ID Y199\_MYCTU  
 AC 007733; ORYJ07;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV1899C.  
 GN RV1899C OR MT1950 OR MTCY180.19.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacteriinae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE=96295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 RP STRAIN-CDC 1551 / Oshkosh;  
 RC Felschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Debey R., Dodson R., Gwin M., Hatt D., Hickey E.,  
 RA Kolonel J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE UPF0189  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: 297193; CAB10035.1; ALT\_TINT.  
 DR EMBL: AE007050; AAK46221.1; -;  
 DR TIGR: MT1950; -;  
 DR TubercuList: RV1899C; -;  
 DR InterPro: IPR002589; Alpp.  
 DR Pfam: PF01661; Alpp; 1.  
 DR SMART: SM00506; Alpp; 1.  
 KW Hypothetical protein; Complete proteome.  
 FT DOMAIN 1 184 UNKNOWN.  
 FT DOMAIN 185 359 UPF0189.  
 FT CONFLICT 76 76 A -> S (IN REF. 2).  
 FT CONFLICT 126 126 C -> S (IN REF. 2).  
 FT CONFLICT 126 126 C -> S (IN REF. 2).  
 SQ SEQUENCE 359 AA; 36936 MW; F73BBB4F55B1426 CRC64;

Query Match 4.2%; Score 187.5; DB 1; Length 359;  
 Best Local Similarity 29.0%; Pred. No. 0.00013;  
 Matches 56; Conservative 32; Mismatches 74; Indels 31; Gaps 5;

QY 118 IELSVKDDLTTHAVDAVYVNAANEDLLHGGLALALVKAGFEIOESKQFARYGVSA 177  
 DB 190 IELVHQAQVTKLELDALINANTRLRHNGVYAAATARAGGPELORESIE-KAPIGL 245

OY 178 GEIIVTGAGRLPCKOIHAV-----GPRMEMDKOGCTGKLQRAIVSILNVYIKNTHIK 232  
Db 246 GEAVETTAGDMPARYVHHAATMELGPTSGEITTAATAATLRKA-----DELGCR 295  
OY 233 TVAIIPAISGIFPFPLNCTKTIVETIRVSLQCKPMSNLKEIHLVSNEDPTVAAPKAAS 292  
Db 296 SLALVAFGTGFGFPLDDAARLWGAVRRRPG-----SLQRVVFAVHGDAAERAFSNA- 349  
OY 293 EFILKSELGOET 305  
Db 350 -----IOAGEDT 356

Search completed: May 13, 2003, 11:42:45  
Job time : 31 secs

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 11:38:38 ; Search time 100 Seconds  
(without alignments)  
1759.642 Million cell updates/sec

Title: US-09-830-762-2  
Perfect score: 4447  
Sequence: 1 MDFSVMAGAAAYNEKSGRIT.....PMRPFADHPWRGFGASGSPVD 854

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID      | Description         |
|------------|--------|-------------|--------|---------|---------------------|
| 1          | 4438   | 99.8        | 854    | 09BZL9  | 09BZL9 homo sapien  |
| 2          | 4220.5 | 94.9        | 819    | 09BZL8  | 09BZL8 homo sapien  |
| 3          | 2453   | 55.2        | 830    | 099LFP9 | 099LFP9 mus musculu |
| 4          | 1747   | 39.3        | 330    | 08TCF3  | 08TCF3 homo sapien  |
| 5          | 610    | 13.7        | 1023   | 09ULF2  | 09ULF2 homo sapien  |
| 6          | 420    | 9.4         | 556    | 09NV60  | 09NV60 homo sapien  |
| 7          | 256    | 5.8         | 419    | 09H9X9  | 09H9X9 homo sapien  |
| 8          | 234.5  | 5.3         | 175    | 08RB30  | 08RB30 thermoaer    |
| 9          | 231    | 5.2         | 298    | 08OCB0  | 08OCB0 oncorhynch   |
| 10         | 230.5  | 5.2         | 424    | 08VDT6  | 08VDT6 mus musculu  |
| 11         | 215    | 4.8         | 183    | 08UDP9  | 08UDP9 pyrococcus   |
| 12         | 211.5  | 4.8         | 1025   | 096K72  | 096K72 homo sapien  |
| 13         | 207    | 4.7         | 672    | 08WV05  | 08WV05 homo sapien  |
| 14         | 205    | 4.6         | 393    | 096CH7  | 096CH7 homo sapien  |
| 15         | 203    | 4.6         | 363    | 09H8R9  | 09H8R9 homo sapien  |
| 16         | 200.5  | 4.5         | 654    | 08R133  | 08R133 mus musculu  |

|    |       |     |      |    |        |                    |
|----|-------|-----|------|----|--------|--------------------|
| 17 | 197   | 4.4 | 606  | 11 | 08R109 | 08R109 mus musculu |
| 18 | 195.5 | 4.4 | 227  | 4  | 08IAJ3 | 08IAJ3 homo sapien |
| 19 | 190   | 4.3 | 243  | 4  | 09B069 | 09B069 homo sapien |
| 20 | 190   | 4.3 | 325  | 4  | 09U986 | 09U986 homo sapien |
| 21 | 189   | 4.3 | 359  | 4  | 09YAP7 | 09YAP7 homo sapien |
| 22 | 186.5 | 4.2 | 175  | 16 | 08RHQ2 | 08RHQ2 fusobacteri |
| 23 | 182   | 4.1 | 195  | 17 | 08RQD0 | 08RQD0 methanosarc |
| 24 | 182   | 4.1 | 243  | 11 | 0922B1 | 0922B1 mus musculu |
| 25 | 178   | 4.0 | 193  | 10 | 0228P6 | 0228P6 arabidopsis |
| 26 | 177   | 4.0 | 193  | 10 | 0949P6 | 0949P6 arabidopsis |
| 27 | 176   | 4.0 | 499  | 12 | 08QUT8 | 08QUT8 infectious  |
| 28 | 171.5 | 3.9 | 182  | 17 | 08WTW2 | 08WTW2 methanopyru |
| 29 | 168   | 3.8 | 369  | 13 | 093326 | 093326 gallus gall |
| 30 | 166   | 3.7 | 369  | 11 | 091VZ2 | 091VZ2 mus musculu |
| 31 | 165.5 | 3.7 | 4383 | 12 | 08V6W7 | 08V6W7 mus musculu |
| 32 | 161.5 | 3.6 | 372  | 4  | 09H8P3 | 09H8P3 bovine coro |
| 33 | 161.5 | 3.6 | 372  | 4  | 09H8P1 | 09H8P1 homo sapien |
| 34 | 160.5 | 3.6 | 2447 | 12 | 09WUD0 | 09WUD0 venezuelan  |
| 35 | 160   | 3.6 | 203  | 5  | 017432 | 017432 caenorhabdi |
| 36 | 159.5 | 3.6 | 4383 | 12 | 091A28 | 091A28 bovine coro |
| 37 | 159.5 | 3.6 | 7094 | 12 | 091A29 | 091A29 bovine coro |
| 38 | 159.5 | 3.6 | 7094 | 12 | 08VA40 | 08VA40 bovine coro |
| 39 | 159.5 | 3.6 | 7094 | 12 | 08VA39 | 08VA39 bovine coro |
| 40 | 157.5 | 3.5 | 1181 | 5  | 09XZ37 | 09XZ37 drosophila  |
| 41 | 157.5 | 3.5 | 1181 | 5  | 09VBP3 | 09VBP3 drosophila  |
| 42 | 156.5 | 3.5 | 372  | 11 | 09QZQ8 | 09QZQ8 mus musculu |
| 43 | 156   | 3.5 | 258  | 4  | 09H610 | 09H610 homo sapien |
| 44 | 156   | 3.5 | 371  | 11 | 009140 | 009140 rattus norv |
| 45 | 156   | 3.5 | 598  | 4  | 09NT13 | 09NT13 homo sapien |

ALIGNMENTS

RESULT 1

ID 09BZL9 PRELIMINARY; PRT; 854 AA.  
AC 09BZL9;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE B aggressive lymphoma long isoform.  
GN BAL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20563954; PubMed=11110709;  
RA Aguilar R.C.T., Yakushiji Y., Kharbanda S., Salgia R., Fletcher J.A.,  
RA Shipp M.A.;  
RT "BAL is a novel risk-related gene in diffuse large B-cell lymphomas  
that enhances cellular migration.";  
RL Blood 96:4328-4334(2000).  
DR EMBL; AF307338; AAK07558.1;  
DR InterPro; IPR002589; A1pp.  
DR Pfam; PF01661; A1pp; 2.  
DR SMART; SM00506; A1pp; 2.  
SO SEQUENCE 854 AA; 96283 MW; E80663868ECCDC03 CRC64;

Query Match 99.8%; Score 4438; DB 4; Length 854;  
Best Local Similarity 99.9%; Pred. No. 4.5e-301;  
Matches 853; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MDFSVMAGAAAYNEKSGITSLIPKVFAPQIPQWKRKNTRECLPYKSGTALGANY 60  
DB 1 MDFSVMAGAAAYNEKSGITSLIPKVFAPQIPQWKRKNTRECLPYKSGTALGANY 60  
OY 61 SMOPTINNDPFLIKNNRQCEVLONKFGCSTLVSVOGNSKSLQVFRKMLTPREL 120  
DB 61 SMOPTINNDPFLIKNNRQCEVLONKFGCSTLVSVOGNSKSLQVFRKMLTPREL 120

```

QY 121 SVWKDLTTAAVDVAVNAANEDLLHGGGLALALVYKAGGFEIOESKOFVARYGVSAGEI 180
D 121 SVWKDLTTAAVDVAVNAANEDLLHGGGLALALVYKAGGFEIOESKOFVARYGVSAGEI 180
QY 181 AVTAGRPLPCQOIITHAVGPRMENDKOGCTGKLORAIVSILNVIYKNTHTKTAIPALS 240
D 181 AVTAGRPLPCQOIITHAVGPRMENDKOGCTGKLORAIVSILNVIYKNTHTKTAIPALS 240
QY 241 SGIFQFPLNLTCTKTIVETIRVSLQGRPMNSNLKEIHLVSNEDPTVAFAKASEFILLKSE 300
D 241 SGIFQFPLNLTCTKTIVETIRVSLQGRPMNSNLKEIHLVSNEDPTVAFAKASEFILLKSE 300
QY 301 LGOETTSFNAAMVNNITLQIVOGHIEWQTAADVIVNSVNPMDITVGVAKSILQOAGVEM 360
D 301 LGOETTSFNAAMVNNITLQIVOGHIEWQTAADVIVNSVNPMDITVGVAKSILQOAGVEM 360
QY 361 KSEFLATKAKOFORSOLVLTGKFNLCFKYIYHVLHMSSEFPKPOLIKHAMKECLEKIEQ 420
D 361 KSEFLATKAKOFORSOLVLTGKFNLCFKYIYHVLHMSSEFPKPOLIKHAMKECLEKIEQ 420
QY 421 NITSISFPALGTGMEIKETAEILFEDEVLTFAKDHVKHQLTVKFVIFPTDLEIYKAFS 480
D 421 NITSISFPALGTGMEIKETAEILFEDEVLTFAKDHVKHQLTVKFVIFPTDLEIYKAFS 480
QY 481 SEMAKRSKMLSNVYSPPOSTREKRENGLEARSAPAINLMGFVNEEMCEAHAWTORILSL 540
D 481 SEMAKRSKMLSNVYSPPOSTREKRENGLEARSAPAINLMGFVNEEMCEAHAWTORILSL 540
QY 541 QNHIIENNHLILYGRKEHDILSOLQTSVSIETIISPGTELEIIGARADILEVYNNI 600
D 541 QNHIIENNHLILYGRKEHDILSOLQTSVSIETIISPGTELEIIGARADILEVYNNI 600
QY 601 EDMCKVQVEEMARKKREGLMSLGOWTIOOQKODEMKENIIFLKCPVPTQELLDOKQ 660
D 601 EDMCKVQVEEMARKKREGLMSLGOWTIOOQKODEMKENIIFLKCPVPTQELLDOKQ 660
QY 661 FEKGLQVLAKEVEDNEVLMAAFORKKKMEKELHROPVSHRLFOQVPOFCVNCVYRGV 720
D 661 FEKGLQVLAKEVEDNEVLMAAFORKKKMEKELHROPVSHRLFOQVPOFCVNCVYRGV 720
QY 721 QRMSTPCDPKYGAGIYFTNKLNLAEKAKISADKLIYFEAEVLTGFCGCHPLNTY 780
D 721 QRMSTPCDPKYGAGIYFTNKLNLAEKAKISADKLIYFEAEVLTGFCGCHPLNTY 780
QY 781 PPLSPGALDGHDSVDVNSSPETFVIFSGMOAIPQVLTCTQEVYOSDYSSGPMRPA 840
D 781 PPLSPGALDGHDSVDVNSSPETFVIFSGMOAIPQVLTCTQEVYOSDYSSGPMRPA 840
QY 841 QHPWRGFASSGSPVD 854
D 841 QHPWRGFASSGSPVD 854

```

## RESULT 2

```

Q9BZL8
ID Q9BZL8 PRELIMINARY: PRT: 819 AA.
AC Q9BZL8:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE B aggressive lymphoma short isoform.
GN BAL.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20563954; PubMed=11110709;
RA Aguilar R.C.T., Yakushiji Y., Kharbada S., Salgia R., Fletcher J.A.,
RT Ship M.A.;
RT "BAL is a novel risk-related gene in diffuse large B-cell lymphomas
RT that enhances cellular migration.";

```

```

RL Blood 96.4328-4334(2000).
DR EMBL: AF307339; AAK07559.1;
DR Interpro: IPR002589; Alpp.
DR Pfam: PF01661; Alpp: 2.
DR SMART: SM00506; Alpp: 2.
SQ SEQUENCE 819 AA; 92211 MW; 32BD9AB99E95837 CRC64;

```

```

Query Match 94.94; Score 4220.5; DB 4; Length 819;
Best Local Similarity 95.88; Pred. No. 6,4e-286;
Matches 818; Conservative 0; Mismatches 1; Indels 35; Gaps 1;

```

```

QY 1 MDFSVMAGAAVYNNKSGRITSLSILFPKVFQAFQFPQMRKNTBCECLPYKSGEALGENTY 60
D 1 MDFSVMAGAAVYNNKSGRITSLSILFPKVFQAFQFPQMRKNTBCECLPYKSGEALGENTY 60
QY 61 SMOPIRHNDFKILKNNEPOLCEVLQKFCISTLVSPVOGNSKSLQVRRKMLTPRIEL 120
D 61 SMOPIRHNDFKILKNNEPOLCEVLQKFCISTLVSPVOGNSKSLQVRRKMLTPRIEL 120
QY 26 SMOPIRHNDFKILKNNEPOLCEVLQKFCISTLVSPVOGNSKSLQVRRKMLTPRIEL 85
D 26 SMOPIRHNDFKILKNNEPOLCEVLQKFCISTLVSPVOGNSKSLQVRRKMLTPRIEL 85
QY 121 SVWKDLTTAAVDVAVNAANEDLLHGGGLALALVYKAGGFEIOESKOFVARYGVSAGEI 180
D 121 SVWKDLTTAAVDVAVNAANEDLLHGGGLALALVYKAGGFEIOESKOFVARYGVSAGEI 180
QY 181 AVTAGRPLPCQOIITHAVGPRMENDKOGCTGKLORAIVSILNVIYKNTHTKTAIPALS 240
D 181 AVTAGRPLPCQOIITHAVGPRMENDKOGCTGKLORAIVSILNVIYKNTHTKTAIPALS 240
QY 241 SGIFQFPLNLTCTKTIVETIRVSLQGRPMNSNLKEIHLVSNEDPTVAFAKASEFILLKSE 300
D 241 SGIFQFPLNLTCTKTIVETIRVSLQGRPMNSNLKEIHLVSNEDPTVAFAKASEFILLKSE 300
QY 301 LGOETTSFNAAMVNNITLQIVOGHIEWQTAADVIVNSVNPMDITVGVAKSILQOAGVEM 360
D 301 LGOETTSFNAAMVNNITLQIVOGHIEWQTAADVIVNSVNPMDITVGVAKSILQOAGVEM 360
QY 361 KSEFLATKAKOFORSOLVLTGKFNLCFKYIYHVLHMSSEFPKPOLIKHAMKECLEKIEQ 420
D 361 KSEFLATKAKOFORSOLVLTGKFNLCFKYIYHVLHMSSEFPKPOLIKHAMKECLEKIEQ 420
QY 421 NITSISFPALGTGMEIKETAEILFEDEVLTFAKDHVKHQLTVKFVIFPTDLEIYKAFS 480
D 421 NITSISFPALGTGMEIKETAEILFEDEVLTFAKDHVKHQLTVKFVIFPTDLEIYKAFS 480
QY 481 SEMAKRSKMLSNVYSPPOSTREKRENGLEARSAPAINLMGFVNEEMCEAHAWTORILSL 540
D 481 SEMAKRSKMLSNVYSPPOSTREKRENGLEARSAPAINLMGFVNEEMCEAHAWTORILSL 540
QY 541 QNHIIENNHLILYGRKEHDILSOLQTSVSIETIISPGTELEIIGARADILEVYNNI 600
D 541 QNHIIENNHLILYGRKEHDILSOLQTSVSIETIISPGTELEIIGARADILEVYNNI 600
QY 601 EDMCKVQVEEMARKKREGLMSLGOWTIOOQKODEMKENIIFLKCPVPTQELLDOKQ 660
D 601 EDMCKVQVEEMARKKREGLMSLGOWTIOOQKODEMKENIIFLKCPVPTQELLDOKQ 660
QY 661 FEKGLQVLAKEVEDNEVLMAAFORKKKMEKELHROPVSHRLFOQVPOFCVNCVYRGV 720
D 661 FEKGLQVLAKEVEDNEVLMAAFORKKKMEKELHROPVSHRLFOQVPOFCVNCVYRGV 720
QY 721 QRMSTPCDPKYGAGIYFTNKLNLAEKAKISADKLIYFEAEVLTGFCGCHPLNTY 780
D 721 QRMSTPCDPKYGAGIYFTNKLNLAEKAKISADKLIYFEAEVLTGFCGCHPLNTY 780
QY 781 PPLSPGALDGHDSVDVNSSPETFVIFSGMOAIPQVLTCTQEVYOSDYSSGPMRPA 840
D 781 PPLSPGALDGHDSVDVNSSPETFVIFSGMOAIPQVLTCTQEVYOSDYSSGPMRPA 840
QY 841 QHPWRGFASSGSPVD 854
D 841 QHPWRGFASSGSPVD 854

```

## RESULT 3

Q99LF9 PRELIMINARY; PRT; 830 AA.

AC Q99LF9; 099LF9; 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, last sequence update)

DE Hypothetical 92.7 kDa protein.

GN BC003281.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RU Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC003281; AAH03281.1; -

DR MGI:1993117; BC003281.

DR InterPro: IPR002589; Alpp.

DR InterPro: IPR001901; SecE.

DR Pfam: PF01661; Alpp; 2.

DR SMART: SM00506; Alpp; 2.

DR PROSITE: PS01067; SECE\_SEC61G; UNKNOWN\_1.

KW Hypothetical protein.

SQ SEQUENCE 830 AA; 92669 MW; 779197039599012 CRC64;

Query Match 55.2%; Score 2453; DB 11; Length 830;

Best Local Similarity 61.4%; Pred. No. 1,6e-162;

Matches 508; Conservative 96; Mismatches 206; Indels 18; Gaps 7;

QY 41 NTECELPYKCSGTCALGENTSWQIPINHNDFILKNNEROLCEVLONKFGCISTLVSPOQ 100

DB 6 DTWAAAEPRPANNLSLEEHQWQIPKHNVEILKSNESOLCEVLONKFGCISTLSPPTL 65

QY 101 EGNKSLQ-VPRKKLTPRIELSVKMDLTHAVDAVVAANMEDLHGGGALALVYKAGF 159

DB 66 AGSSPAPQVRFKRLTLPGIELSVKMDLTHAVDAVVAANMEDLHGGGALALVYKAGF 125

QY 160 EIOESKOFARYKVSAGELAVTAGRLPCQIILHAVGPRMMDQOGCTGKLORAVS 219

DB 126 EIOESKRIIANVKGISVGAITAGRLPCQIILHAVGPRMMDQOGCTGKLORAVS 185

QY 220 ILNVYIKNNHIKVAIPALSSGIFOPPLNCTIVETIVTSLOQYPRMNLKEIHLS 279

DB 186 ILDVYTDIKIKVAIPALSSGIFOPPLNCTIVETIVTSLOQYPRMNLKEIHLS 245

QY 280 NEDPTVAFAKASEPILGK--SELGOETPP--SFNAMVAVNLTLOIYQGIEMQTADV 334

DB 246 NEDPTVASFKSASESILIGRLSSMGPEPTDPAITMLRIGGLTLQIYQGIEMQTADV 305

QY 335 VNSVNPDIIVGPAKSILOQAVEMKSEFLATRAKQFORSOLVLYTKGFNLFCRYIYHV 394

DB 306 VNSVYMDDFKSGRAVQASILQAVEMKSEFLATRAKQFORSOLVLYTKGFNLFCRYIYHV 363

QY 395 LHMSEFPKPOLKHAMKECLEKTEONTSTSPALGAGNNEIKKETAALTLPEVLTFA 454

DB 364 AMHSQINKYQILKQAMKSCLEKCLPDINSTSPALGAGNNEIKKETAALTLPEVLTFA 423

QY 455 KDHQKHLDTLVFVFPDLLEYKAFSSPMARSKMLSLNYS--VPSTREEKRENGLE 511

DB 424 KEHEKTLTVKIVIFPVDVETIKIFYAEMTKRSNELNSGSGALALQWSSGEOGRGLE 483

QY 512 ARSPAINLMGFVEMTEFAHAMIORILSLQNHIIENNHLIYLGKREHDLISOLQKTSV 571

DB 484 AGSPAINLMGVKMECAQEMIERLLVSLDHIIENNHLIYLGKREHDLISOLQKTSV 543

QY 572 SITEILSPGRTELEEGARADLIEVYNIEDMLCKVQDEMAKKEKRGKWRSLQWTIOQ 631

DB 544 SITEIVSPRTALIEKQADLIDAVMRIECLDVQDEVAGKRRKMLMSLSGGOTNOE 603

QY 632 KTOEMKKNITFLKCPVPTOELLDOKKOFKCGLOVYKVKRINDEYVMAAFORKKKME 691

DB 604 KL-DKMESTYFQRPASLTQELDRKKQFCKGILMVVQVQIDNKVLLAFLQEKKKME 662

QY 692 EKLHROPVSHRLFOOVYOFNCNVYCRVGFQRMVSTPCDPKYGAGIYFTKMLAEKAKK 751

DB 663 ERFPKSGSQRLLQOVYHQCNCVYCRVGFQRMVSTPCDPKYGAGIYFTKMLAEKAKK 722

QY 752 ISADKLIVFEAEVLGFCQGHPLNIVPPPLSPGALIDGSDVVDVNSPETTFVIFSGM 811

DB 723 TSTSDKLIVFEAEVLGFCQGHPLNIVPPPLSPGALIDGSDVVDVNSPETTFVIFSGM 782

QY 812 QAIQOYLMTCTOEY-----VQSDYSSGPMRPAQHPRMGFASGSFV 853

DB 783 QAMPYLMTCTQDRTFSQHPMWSGYSYSGPMVSSLSQWEMVINGSSV 830

RESULT 4

QY 08TCB3 PRELIMINARY; PRT; 330 AA.

AC 08TCB3; 08TCB3; 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)

DE Hypothetical 37.9 kDa protein (Fragment).

GN DKFZP6111617.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA TISSUE-AIRGADALA;

RA Ansoorge W., Winkner U., Nemes H.W., Well B., Wiemann S.;

RU Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL713679; CAD28483.1; -

KW Hypothetical protein.

FT NON\_TER

SQ SEQUENCE 330 AA; 37944 MW; 7E6DB2BD3E16B4C CRC64;

Query Match 39.3%; Score 1747; DB 4; Length 330;

Best Local Similarity 100.0%; Pred. No. 7.2e-114;

Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 525 EEMEYEAHAMIORILSLQNHIIENNHLIYLGKREHDLISOLQKTSVSTIEILSPGRT 584

DB 1 EEMEYEAHAMIORILSLQNHIIENNHLIYLGKREHDLISOLQKTSVSTIEILSPGRT 60

QY 585 EIEGARADLIEVYNIEDMLCKVQDEMAKKEKRLMSLGWTIOQKTODEMKNITFL 644

DB 61 EIEGARADLIEVYNIEDMLCKVQDEMAKKEKRLMSLGWTIOQKTODEMKNITFL 120

QY 645 KCPVPTOELLDOKKQFCKGLOVLYKVEKIDNEVYMAAFORKKKMEKRLHROPVSHRL 704

DB 121 KCPVPTOELLDOKKQFCKGLOVLYKVEKIDNEVYMAAFORKKKMEKRLHROPVSHRL 180

QY 705 QOVPYOFNCNVYCRVGFQRMVSTPCDPKYGAGIYFTKMLAEKAKKISADKLIVFEA 764

DB 181 QOVPYOFNCNVYCRVGFQRMVSTPCDPKYGAGIYFTKMLAEKAKKISADKLIVFEA 240

QY 765 EVLTGFCQGHPLNIVPPPLSPGALIDGSDVVDVNSPETTFVIFSGMALPOYLMTCTOE 824

DB 241 EVLTGFCQGHPLNIVPPPLSPGALIDGSDVVDVNSPETTFVIFSGMALPOYLMTCTOE 300

QY 825 YVQSDYSSGPMRPAQHPRMGFASGSFVD 854

DB 301 YVQSDYSSGPMRPAQHPRMGFASGSFVD 330

RESULT 5

QY 09ULF2 PRELIMINARY; PRT; 1023 AA.

AC 09ULF2; 09ULF2; 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)



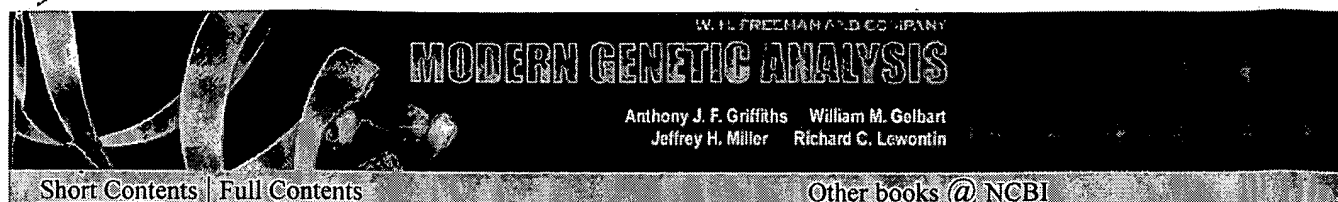


0y 635 DEMENITIELFCBPVPNOELLDDKKOEKGLQ--VLKVEIDNEVLAAPFORKKME 6922  
::: ::: :  
Db 84 DDMKSLLMLVPLPFGSGEGHDIKEERKTLLNNTILTIERYQNDSDLMKSYOIRKLLEE 1433  
:  
0y 693 KLRHPVSHRLFOOVPIQFCNVYCRVGFORMYSTPCDPKGAGIYFTFKNLKILAERAAKT 7522  
:  
Db 144 KNKHPTNEKKLLEFHOTSSDSTIQINHHENRYSYAQTGHGAALONGSEYFVNSSYSARGSKA 203



[illegible]





## Navigation

### About this book

#### 11.

#### Applications of Recombinant DNA Technology

#### Introduction

#### In Vitro

#### Mutagenesis

#### → RFLP Mapping

#### Reverse Genetics

#### Expressing Eukaryotic Genes in Bacteria

#### Recombinant DNA Technology in Eukaryotes

#### Gene Therapy

#### Using Recombinant DNA to Detect Disease Alleles Directly

#### Summary

#### Concept Map

#### Solved Problem

#### Solution

#### Solved Problem

#### Problems

## Figures

### Figure 11-2.

#### Mapping a gene...

### Figure 11-3.

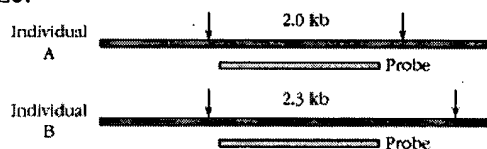
#### The detection and...

## Search

## Modern Genetic Analysis → 11. Applications of Recombinant DNA Technology

### RFLP Mapping

In Chapter 10, we learned that, if a cloned DNA fragment is used as a probe of genomic DNA that has been cut with a restriction enzyme, then the probe will bind to one or more genomic fragments. For example, if the restriction enzyme used does not cut within the chromosomal region encompassed by the cloned fragment, then the probe should bind to one fragment flanked by restriction sites on each side. Since the DNA of chromosomes within a species is generally homologous, it might be expected that a constant-sized genomic fragment will be bound in all individuals. However, when probes are used in this way, the bound fragments are often found to be of different sizes in different individuals. The explanation is that a given restriction site is not always found in all individuals. The absence of a site is usually caused by a single nucleotide difference that is most likely biologically neutral. Hence, for example, if a probe binds a 2-kb fragment in individual A of a haploid species and it binds a fragment of 2.3 kb in individual B, the reason is usually that one of the sites that flanked the 2-kb fragment is missing in B, and the next site is 0.3 kb away, making the hybridized fragment 2.3 kb in size.



The presence and absence of the restriction site can be treated as two alleles that can be thought of as + and - alleles. The presence of the + in some individuals in the population and the absence (-) in others generates a **restriction fragment length polymorphism, or RFLP**. (In the case just discussed, there was a dimorphism—two “morphs,” one short and one long.) Geneticists were surprised to discover that RFLPs are quite common in populations and that a large proportion of probes will detect one. RFLPs are identified by a rather hit-or-miss method of hybridizing panels of randomly cloned genomic fragments to genomic restriction digests of several different individuals in a family or a population. Because RFLPs are a relatively common type of variation in nature, this method succeeds in finding RFLPs in most cases.

The significance of RFLPs is threefold. First, if an individual is heterozygous for two morphs of an RFLP, this heterozygous “locus” can be used as a marker in chromosomal mapping. Although at first the locus of the RFLP is not necessarily known, as more and more RFLPs are found, they can be mapped in relation to gene loci and in relation to other RFLP loci, and their positions gradually saturate the genetic map. The RFLPs are not biologically significant in most cases, but they can be used to map interesting genes and act as positions from which these

**THIS PAGE BLANK (USPTO)**

Go

- ☒ This book  
☐ All books  
☐ PubMed

genes can be cloned by positional cloning.

Second, in an extension of mapping analysis, RFLP alleles (morphs) can be used as diagnostic tools. For example, in a family with a record of a certain disease, if it can be established that the people who have the disease also carry a specific allele of an RFLP, then this fact suggests not only that the RFLP locus is linked to the disease gene locus, but furthermore that the specific RFLP allele is in cis arrangement with the disease allele. Hence the RFLP allele becomes a diagnostic marker for the disease, and this information can be used in genetic counseling.

Third, RFLPs can be used to measure genetic divergence between different populations or related species. The restriction-site difference is effectively a DNA difference, so a measure of the total number of RFLP differences represents a measure of genetic difference. Hence RFLPs are important in studies of evolution.

RFLP mapping is often performed on a defined set of strains or individuals that become “standards” for mapping that species. For example, in the fungus *Neurospora*, two wild-type strains, Oak Ridge and Mauriceville, are known to show many RFLP differences, so these strains have become standards used in RFLP mapping. The RFLPs can be mapped relative to one another or to genes of known phenotypic expression. For example, let *ad* stand for an allele for adenine requirement, and 1 and 2 stand for RFLP loci with either the Oak Ridge (OR) or Mauriceville (M) “alleles.” A cross can be made of the type  $ad \cdot 1^{OR} \cdot 2^{OR} \times ad^{+} \cdot 1^{M} \cdot 2^{M}$ . Progeny are tested for all three loci. Adenine requirement is tested by inoculating strains on medium lacking adenine, and the RFLP alleles are tested by probing with the relevant probes. Recombinant frequencies are calculated in the usual way. Most mutants in *Neurospora* have been induced in Oak Ridge wild-type strains, so it is a simple matter to map the mutant alleles to RFLPs simply by crossing the mutant Oak Ridge strain to the wild-type Mauriceville strain. An example of mapping a phenotypic mutant by using RFLP markers is shown in [Figure 11-2](#).

Similar standard strains have been established in other organisms. An analogous approach has been used in human genome mapping by collecting DNA from a defined set of individuals in 61 families with an average of eight children per family and making this DNA available throughout the world to provide a standard for RFLP mapping.

[Figure 11-3](#) shows an example of linkage of a human disease allele to an RFLP locus and the potential for using this information in diagnostics. Because of the close linkage, future generations of persons showing the RFLP morph 1 can be predicted to have a high chance of inheriting the disease allele *D*. This sort of predictive power can be used in prenatal diagnoses of the genotypes of fetuses, with the use of amniocentesis or chorionic villus sampling (considered later in this chapter).

**THIS PAGE BLANK (USPTO)**



**MESSAGE**

**RFLPs provide useful molecular marker loci for chromosome mapping and for diagnosis of human disease alleles.**

It is worth comparing the process of making a restriction map (*restriction mapping*, pages 327–329) with the process of *RFLP mapping*. Restriction maps are based on *physical* analysis of DNA, whereas RFLP maps are based on *recombination* analysis of matings. Note also that restriction mapping is based on restriction sites with no variation, whereas RFLP mapping is based on restriction-site variation between homologous chromosomes. Most restriction maps are short-range (fine-scale) maps, although long-range maps can be constructed with rare-cutting restriction enzymes. In contrast, RFLP mapping generally produces long-range (coarse-scale) maps. RFLP mapping of whole genomes will be covered in detail in Chapter 12. ♦ [TOP](#)

© 1999 by W. H. Freeman and Company. All rights reserved.

**THIS PAGE BLANK (USPTO)**